

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 June 2003 (19.06.2003)

PCT

(10) International Publication Number  
**WO 03/050253 A2**

(51) International Patent Classification<sup>7</sup>: C12N

(21) International Application Number: PCT/US02/39133

(22) International Filing Date: 4 December 2002 (04.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/340,747	7 December 2001 (07.12.2001)	US
60/342,761	20 December 2001 (20.12.2001)	US
60/349,705	15 January 2002 (15.01.2002)	US
60/354,764	6 February 2002 (06.02.2002)	US
60/356,216	12 February 2002 (12.02.2002)	US

CA 95118 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **TRAN, Bao** [US/US]; 750 Salberg Avenue, Santa Clara, CA 95051 (US). **SPRAGUE, William, W.** [US/US]; 611 13th Street #C, Sacramento, CA 95814 (US). **LEE, Soo, Yeun** [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). **KHARE, Reena** [US/US]; 12650 Orella Court, Saratoga, CA 95070 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **BHATIA, Umesh** [US/US]; 5212 Union Avenue, San Jose, CA 95124 (US). **BURRILL, John, D.** [US/US]; 2218 Brewster Avenue, Redwood City, CA 94062 (US). **BLAKE, Julie, J.** [US/US]; 3818 Pacheco Street, San Francisco, CA 94116 (US). **HO, Anne** [FR/US]; 1279 Poplar Avenue #114, Sunnyvale, CA 94086 (US). **ZHENG, Wenjin** [CN/US]; 9 Sutter Creek Lane, Mountain View, CA 94043 (US).

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue #71, Palo Alto, CA 94303 (US). **KABLE, Amy, E.** [US/US]; 2345 Polk Street #4, San Francisco, CA 94109 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **MARQUIS, Joseph, P.** [US/US]; 4428 Lazy Lane, San Jose, CA 95135 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **GORVAD, Ann, E.** [US/US]; 369 Marie Common, Livermore, CA 94550 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive #117, Castro Valley, CA 94546 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **LEE, Sally** [US/US]; 3643 26th Street, San Francisco, CA 94110 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose,

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

(57) Abstract: Various embodiments of the invention provide human proteins associated with cell growth, differentiation, and death (CGDD) and polynucleotides which identify and encode CGDD. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of CGDD.

**PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH****TECHNICAL FIELD**

The invention relates to novel nucleic acids, proteins associated with cell growth,  
5 differentiation, and death encoded by these nucleic acids, and to the use of these nucleic acids and  
proteins in the diagnosis, treatment, and prevention of cell proliferative disorders including cancer,  
developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive  
disorders, and disorders of the placenta. The invention also relates to the assessment of the effects of  
exogenous compounds on the expression of nucleic acids and proteins associated with cell growth,  
10 differentiation, and death.

**BACKGROUND OF THE INVENTION**

Human growth and development requires the spatial and temporal regulation of cell  
differentiation, cell proliferation, and apoptosis. These processes coordinately control reproduction,  
15 aging, embryogenesis, morphogenesis, organogenesis, and tissue repair and maintenance. At the  
cellular level, growth and development is governed by the cell's decision to enter into or exit from  
the cell division cycle and by the cell's commitment to a terminally differentiated state. These  
decisions are made by the cell in response to extracellular signals and other environmental cues it  
receives. The following discussion focuses on the molecular mechanisms of cell division,  
20 embryogenesis, cell differentiation and proliferation, and apoptosis, as well as disease states such as  
cancer which can result from disruption of these mechanisms.

**Cell Cycle**

Cell division is the fundamental process by which all living things grow and reproduce. In  
unicellular organisms such as yeast and bacteria, each cell division doubles the number of organisms.  
25 In multicellular species many rounds of cell division are required to replace cells lost by wear or by  
programmed cell death, and for cell differentiation to produce a new tissue or organ. Progression  
through the cell cycle is governed by the intricate interactions of protein complexes. This regulation  
depends upon the appropriate expression of proteins which control cell cycle progression in response  
to extracellular signals, such as growth factors and other mitogens, and intracellular cues, such as  
30 DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle  
progression fall into several categories, including cyclins, cyclin-dependent protein kinases, growth  
factors and their receptors, second messenger and signal transduction proteins, oncogene products,  
and tumor-suppressor proteins.

Regulated progression of the cell cycle depends on the integration of growth control  
35 pathways with the basic cell cycle machinery. Cell cycle regulators have been identified by selecting

for human and yeast cDNAs that block or activate cell cycle arrest signals in the yeast mating pheromone pathway when they are overexpressed. Known regulators include human CPR (cell cycle progression restoration) genes, such as CPR8 and CPR2, and yeast CDC (cell division control) genes, including CDC91, that block the arrest signals. The CPR genes express a variety of proteins  
5 including cyclins, tumor suppressor binding proteins, chaperones, transcription factors, translation factors, and RNA-binding proteins (Edwards, M.C. et al.(1997) Genetics 147:1063-1076).

Several cell cycle transitions, including the entry and exit of a cell from mitosis, are dependent upon the activation and inhibition of cyclin-dependent kinases (Cdks). The Cdks are composed of a kinase subunit, Cdk, and an activating subunit, cyclin, in a complex that is subject to  
10 many levels of regulation. There appears to be a single Cdk in Saccharomyces cerevisiae and Saccharomyces pombe whereas mammals have a variety of specialized Cdks. Cyclins act by binding to and activating cyclin-dependent protein kinases which then phosphorylate and activate selected proteins involved in the mitotic process. The Cdk-cyclin complex is both positively and negatively regulated by phosphorylation, and by targeted degradation involving molecules such as CDC4 and  
15 CDC53. In addition, Cdks are further regulated by binding to inhibitors and other proteins such as Suc1 that modify their specificity or accessibility to regulators (Patra, D. and W.G. Dunphy (1996) Genes Dev. 10:1503-1515; and Mathias, N. et al. (1996) Mol. Cell Biol. 16:6634-6643).

### **Reproduction**

The male and female reproductive systems are complex and involve many aspects of growth  
20 and development. The anatomy and physiology of the male and female reproductive systems are reviewed in (Guyton, A.C. (1991) Textbook of Medical Physiology, W.B. Saunders Co., Philadelphia PA, pp. 899-928).

The male reproductive system includes the process of spermatogenesis, in which the sperm are formed, and male reproductive functions are regulated by various hormones and their effects on  
25 accessory sexual organs, cellular metabolism, growth, and other bodily functions.

Spermatogenesis begins at puberty as a result of stimulation by gonadotropic hormones released from the anterior pituitary. Immature sperm (spermatogonia) undergo several mitotic cell divisions before undergoing meiosis and full maturation. The testes secrete several male sex hormones, the most abundant being testosterone, that is essential for growth and division of the  
30 immature sperm, and for the masculine characteristics of the male body. Three other male sex hormones, gonadotropin-releasing hormone (GnRH), luteinizing hormone (LH), and follicle-stimulating hormone (FSH) control sexual function.

The uterus, ovaries, fallopian tubes, vagina, and breasts comprise the female reproductive system. The ovaries and uterus are the source of ova and the location of fetal development,  
35 respectively. The fallopian tubes and vagina are accessory organs attached to the top and bottom of

the uterus, respectively. Both the uterus and ovaries have additional roles in the development and loss of reproductive capability during a female's lifetime. The primary role of the breasts is lactation. Multiple endocrine signals from the ovaries, uterus, pituitary, hypothalamus, adrenal glands, and other tissues coordinate reproduction and lactation. These signals vary during the monthly  
5 menstruation cycle and during the female's lifetime. Similarly, the sensitivity of reproductive organs to these endocrine signals varies during the female's lifetime.

A combination of positive and negative feedback to the ovaries, pituitary and hypothalamus glands controls physiologic changes during the monthly ovulation and endometrial cycles. The anterior pituitary secretes two major gonadotropin hormones, follicle-stimulating hormone (FSH) and  
10 luteinizing hormone (LH), regulated by negative feedback of steroids, most notably by ovarian estradiol. If fertilization does not occur, estrogen and progesterone levels decrease. This sudden reduction of the ovarian hormones leads to menstruation, the desquamation of the endometrium.

Hormones further govern all the steps of pregnancy, parturition, lactation, and menopause. During pregnancy large quantities of human chorionic gonadotropin (hCG), estrogens, progesterone,  
15 and human chorionic somatomammotropin (hCS) are formed by the placenta. hCG, a glycoprotein similar to luteinizing hormone, stimulates the corpus luteum to continue producing more progesterone and estrogens, rather than to involute as occurs if the ovum is not fertilized. hCS is similar to growth hormone and is crucial for fetal nutrition.

The female breast also matures during pregnancy. Large amounts of estrogen secreted by the  
20 placenta trigger growth and branching of the breast milk ductal system while lactation is initiated by the secretion of prolactin by the pituitary gland.

Parturition involves several hormonal changes that increase uterine contractility toward the end of pregnancy, as follows. The levels of estrogens increase more than those of progesterone. Oxytocin is secreted by the neurohypophysis. Concomitantly, uterine sensitivity to oxytocin  
25 increases. The fetus itself secretes oxytocin, cortisol (from adrenal glands), and prostaglandins.

Menopause occurs when most of the ovarian follicles have degenerated. The ovary then produces less estradiol, reducing the negative feedback on the pituitary and hypothalamus glands. Mean levels of circulating FSH and LH increase, even as ovulatory cycles continue. Therefore, the ovary is less responsive to gonadotropins, and there is an increase in the time between menstrual  
30 cycles. Consequently, menstrual bleeding ceases and reproductive capability ends.

#### **Cell Differentiation and Proliferation**

Tissue growth involves complex and ordered patterns of cell proliferation, cell differentiation, and apoptosis. Cell proliferation must be regulated to maintain both the number of cells and their spatial organization. This regulation depends upon the appropriate expression of  
35 proteins which control cell cycle progression in response to extracellular signals, such as growth



factors and other mitogens, and intracellular cues, such as DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including growth factors and their receptors, second messenger and signal transduction proteins, oncogene products, tumor-suppressor proteins, and mitosis-promoting factors.

5           Growth factors were originally described as serum factors required to promote cell proliferation. Most growth factors are large, secreted polypeptides that act on cells in their local environment. Growth factors bind to and activate specific cell surface receptors and initiate intracellular signal transduction cascades. Many growth factor receptors are classified as receptor tyrosine kinases which undergo autophosphorylation upon ligand binding. Autophosphorylation  
10 enables the receptor to interact with signal transduction proteins characterized by the presence of SH2 or SH3 domains (Src homology regions 2 or 3). These proteins then modulate the activity state of small G-proteins, such as Ras, Rab, and Rho, along with GTPase activating proteins (GAPs), guanine nucleotide releasing proteins (GNRPs), and other guanine nucleotide exchange factors. Small G proteins act as molecular switches that activate other downstream events, such as mitogen-activated  
15 protein kinase (MAP kinase) cascades. MAP kinases ultimately activate transcription of mitosis-promoting genes.

In addition to growth factors, small signaling peptides and hormones also influence cell proliferation. These molecules bind primarily to another class of receptor, the trimeric G-protein coupled receptor (GPCR), found predominantly on the surface of immune, neuronal and  
20 neuroendocrine cells. Upon ligand binding, the GPCR activates a trimeric G protein which in turn triggers increased levels of intracellular second messengers such as phospholipase C, Ca<sup>2+</sup>, and cyclic AMP. Most GPCR-mediated signaling pathways indirectly promote cell proliferation by causing the secretion or breakdown of other signaling molecules that have direct mitogenic effects. These signaling cascades often involve activation of kinases and phosphatases. Some growth factors,  
25 such as some members of the transforming growth factor beta (TGF- $\beta$ ) family, act on some cells to stimulate cell proliferation and on other cells to inhibit it. Growth factors may also stimulate a cell at one concentration and inhibit the same cell at another concentration. Most growth factors also have a multitude of other actions besides the regulation of cell growth and division: they can control the proliferation, survival, differentiation, migration, or function of cells depending on the circumstance. For example, the tumor necrosis factor/nerve growth factor (TNF/NGF) family can activate or inhibit  
30 cell death, as well as regulate proliferation and differentiation. The cell response depends on the type of cell, its stage of differentiation and transformation status, which surface receptors are stimulated, and the types of stimuli acting on the cell (Smith, A. et al. (1994) Cell 76:959-962; and Nocentini, G. et al. (1997) Proc. Natl. Acad. Sci. USA 94:6216-6221).

35           Neighboring cells in a tissue compete for growth factors, and when provided with "unlimited"

quantities in a perfused system will grow to even higher cell densities before reaching density-dependent inhibition of cell division. Cells often demonstrate an anchorage dependence of cell division as well. This anchorage dependence may be associated with the formation of focal contacts linking the cytoskeleton with the extracellular matrix (ECM). The expression of ECM components  
5 can be stimulated by growth factors. For example, TGF- $\beta$  stimulates fibroblasts to produce a variety of ECM proteins, including fibronectin, collagen, and tenascin (Pearson, C.A. et al. (1988) EMBO J. 7:2677-2981). In fact, for some cell types specific ECM molecules, such as laminin or fibronectin, may act as growth factors. Tenascin-C and -R, expressed in developing and lesioned neural tissue, provide stimulatory/anti-adhesive or inhibitory properties, respectively, for axonal growth (Faissner,  
10 A. (1997) Cell Tissue Res. 290:331-341).

Cancers are associated with the activation of oncogenes which are derived from normal cellular genes. These oncogenes encode oncoproteins which convert normal cells into malignant cells. Some oncoproteins are mutant isoforms of the normal protein, and other oncoproteins are abnormally expressed with respect to location or amount of expression. The latter category of  
15 oncoprotein causes cancer by altering transcriptional control of cell proliferation. Five classes of oncoproteins are known to affect cell cycle controls. These classes include growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. Viral oncogenes are integrated into the human genome after infection of human cells by certain viruses. Examples of viral oncogenes include v-src, v-abl, and v-fps. Many cases related to  
20 the overexpression of proteins associated with tumors and metastasis have been reported. The Mta1 gene has been cloned in mice, in both cell lines and tissues representing metastatic tumors (Simpson, A. et al. (2001) Gene 273:29-39). Expression of the melanoma antigen-encoding gene (MAGE) family of proteins has also been detected in many tumors. GAC1, a new member of the leucine-rich repeat superfamily, is amplified and overexpressed in malignant gliomas (Almeida, A. et al. (1998)  
25 Oncogene 16:2997-3002).

Many oncogenes have been identified and characterized. These include sis, erbA, erbB, her-2, mutated G<sub>s</sub>, src, abl, ras, crk, jun, fos, myc, and mutated tumor-suppressor genes such as RB, p53, mdm2, Cip1, p16, and cyclin D. Transformation of normal genes to oncogenes may also occur by chromosomal translocation. The Philadelphia chromosome, characteristic of chronic myeloid  
30 leukemia and a subset of acute lymphoblastic leukemias, results from a reciprocal translocation between chromosomes 9 and 22 that moves a truncated portion of the proto-oncogene c-abl to the breakpoint cluster region (bcr) on chromosome 22.

Tumor-suppressor genes are involved in regulating cell proliferation. Mutations which cause reduced or loss of function in tumor-suppressor genes result in uncontrolled cell proliferation. For  
35 example, the retinoblastoma gene product (RB), in a non-phosphorylated state, binds several early-

response genes and suppresses their transcription, thus blocking cell division. Phosphorylation of RB causes it to dissociate from the genes, releasing the suppression, and allowing cell division to proceed.

SEB (SET-binding protein) is a novel nuclear protein that interacts in a yeast two-hybrid  
5 system and in human cells with SET, the translocation breakpoint-encoded protein in acute  
undifferentiated leukemia. SEB also has an oncoprotein Ski homologous region, six PEST sequences  
and three sequential PPLPPPP repeats at the C-terminus. SEB mRNA is expressed ubiquitously in  
all examined human adult tissues and cells. SET has been mapped to chromosome 18q21.1. This  
region also contains tumor suppressor genes associated with deletions in cancer and leukemia  
10 (Minakuchi, M. et al. (2001) *Eur. J. Biochem.* 268:1340-1351).

Angiogenesis is the process by which new capillaries are formed by sprouting from  
preexisting vessels. It is a vital function for the growth of normal tissues during embryogenesis as  
well as for the pathological growth of tumors. Pathological proliferation of cancer cells will not  
result in a proportional increase in mass without access to the blood circulation. Tumors form their  
15 own circulatory system by upregulating angiogenic stimulators and by downregulation angiogenesis  
inhibitors. The inability of metastases to induce an angiogenic response results in a dormant  
phenotype. Angiostatin is a circulating inhibitor of angiogenesis. In vitro, it inhibits endothelial cell  
migration, proliferation, and tube formation, and induces apoptosis in a cell type-specific manner.  
Angiomotin is an angiostatin-binding peptide that mediates angiostatin inhibition of migration and  
20 tube formation of endothelial cells. Angiomotin is expressed in the endothelial cells of capillaries  
and the larger vessels of the human placenta. Angiostatin inhibits cell migration by interfering with  
angiomotin activity in endothelial cells (Trojanovsky, B. et al. (2001) *J. Cell Biol.* 152:1247-1254).

Nucleolar protein p120 is a proliferation-associated antigen expressed by cells in early G1  
phase, identified by the monoclonal antibody FB. It is very cancer specific. In fact, quantitative  
25 immunohistochemical analysis of p120 protein is an easy and reliable method for the assessment of  
clinical outcome and the definition of risk groups in oral carcinoma (Ventura, L. et al. (1999)  
*Anticancer Res.* 19:1423-1426). P120 contains a basic domain, an acidic domain, a hydrophobic and  
methionine-rich domain, and a domain rich in cysteine and proline residues (Fonagy, A. et al. (1989)  
*Cancer Commun.* 1:243-251). This protein is expressed in early G1 and has not been detected in  
30 benign tumors and most normal resting tissues. Sato et al. show that the expression level of p120 in  
tumor tissues can be used as an independent and powerful prognostic marker for resected lung  
adenocarcinoma (Sato, G. et al. (1999) *J. Clin. Oncol.* 17:2721-2727).

The human LGI1 gene is a leucine-rich, repeat-containing gene that was cloned from the  
t(10;19) breakpoint of the T98G glioblastoma cell line. The LGI1 gene maps to 10q24, a region of  
35 peak LOH in malignant gliomas, and is inactivated during the transition from low to high-grade brain

tumors. The mouse *lgil* gene is 97% homologous to the human gene at the amino acid level and 91% homologous at the nucleotide level. *LGI1* contains 8 exons, where each of the four leucine-rich repeat units is contained in an individual 72-bp exon. The cysteine-rich regions flanking the LRR and the single trans-membrane domain do not occupy individual exons (Somerville, R.P. et al. (2000)

5 Mamm. Genome 11:622-627).

Details of the cell division cycle may vary, but the basic process consists of three principle events. The first event, interphase, involves preparations for cell division, replication of the DNA, and production of essential proteins. In the second event, mitosis, the nuclear material is divided and separates to opposite sides of the cell. The final event, cytokinesis, is division and fission of the cell  
10 cytoplasm. The sequence and timing of cell cycle transitions is under the control of the cell cycle regulation system which controls the process by positive or negative regulatory circuits at various check points.

Mitosis marks the end of interphase and concludes with the onset of cytokinesis. There are four stages in mitosis, occurring in the following order: prophase, metaphase, anaphase and  
15 telophase. Prophase includes the formation of bi-polar mitotic spindles, composed of microtubules and associated proteins such as dynein, which originate from polar mitotic centers. During metaphase, the nuclear material condenses and develops kinetochore fibers which aid in its physical attachment to the mitotic spindles. The ensuing movement of the nuclear material to opposite poles along the mitotic spindles occurs during anaphase. Telophase includes the disappearance of the  
20 mitotic spindles and kinetochore fibers from the nuclear material. Mitosis depends on the interaction of numerous proteins. For example, centromere-associated proteins such as CENP-A, -B, and -C, play structural roles in kinetochore formation and assembly (Saffery, R. et al. (2000) Human Mol. Gen. 9:175-185).

During the M phase of eukaryotic cell cycling, structural rearrangements occur ensuring  
25 appropriate distribution of cellular components between daughter cells. Breakdown of interphase structures into smaller subunits is common. The nuclear envelope breaks into vesicles, and nuclear lamins are disassembled. Subsequent phosphorylation of these lamins occurs and is maintained until telophase, at which time the nuclear lamina structure is reformed. cDNAs responsible for encoding M phase phosphorylation (MPPs) are components of U3 small nucleolar ribonucleoprotein (snoRNP),  
30 and relocalize to the nucleolus once mitosis is complete (Westendorf, J.M. et al. (1998) J. Biol. Chem. 9:437-449). U3 snoRNPs are essential mediators of RNA processing events.

Proteins involved in the regulation of cellular processes such as mitosis include the Ser/Thr-protein phosphatases type 1 (PP-1). PP-1s act by dephosphorylation of key proteins involved in the metaphase-anaphase transition. The gene *PP1R7* encodes the regulatory polypeptide *sds22*, having at  
35 least six splice variants (Ceulemans, H. et al. (1999) Eur. J. Biochem. 262:36-42). *Sds22* modulates

the activity of the catalytic subunit of PP-1s, and enhances the PP-1-dependent dephosphorylation of mitotic substrates.

Cell cycle regulatory proteins play an important role in cell proliferation and cancer. For example, failures in the proper execution and timing of cell cycle events can lead to chromosome  
5 segregation defects resulting in aneuploidy or polyploidy. This genomic instability is characteristic of transformed cells (Luca, F.C. and M. Winey (1998) Mol. Biol. Cell. 9:29-46). A recently identified protein, mMOB1, is the mammalian homolog of yeast MOB1, an essential yeast gene required for completion of mitosis and maintenance of ploidy. The mammalian mMOB1 is a member of protein complexes including protein phosphatase 2A (PP2A), and its phosphorylation appears to be  
10 regulated by PP2A (Moreno, C.S. et al. (2001) J. Biol. Chem. 276:24253-24260). PP2A has been implicated in the development of human cancers, including lung and colon cancers and leukemias.

Cell cycle regulation involves numerous proteins interacting in a sequential manner. The eukaryotic cell cycle consists of several highly controlled events whose precise order ensures ' successful DNA replication and cell division. Cells maintain the order of these events by making  
15 later events dependent on the successful completion of earlier events. This dependency is enforced by cellular mechanisms called checkpoints. Examples of additional cell cycle regulatory proteins include the histone deacetylases (HDACs). HDACs are involved in cell cycle regulation, and modulate chromatin structure. Human HDAC1 has been found to interact *in vitro* with the human Hus1 gene product, whose *Schizosaccharomyces pombe* homolog has been implicated in G<sub>2</sub>/M  
20 checkpoint control (Cai, R.L. et al. (2000) J. Biol. Chem. 275:27909-27916).

DNA damage (G<sub>2</sub>) and DNA replication (S-phase) checkpoints arrest eukaryotic cells at the G<sub>2</sub>/M transition. This arrest provides time for DNA repair or DNA replication to occur before entry into mitosis. Thus, the G<sub>2</sub>/M checkpoint ensures that mitosis only occurs upon completion of DNA replication and in the absence of chromosomal damage. The Hus1 gene of *Schizosaccharomyces*  
25 *pombe* is a cell cycle checkpoint gene, as are the rad family of genes (e.g., rad1 and rad9) (Volkmer, E. and L.M. Karnitz (1999) J. Biol. Chem. 274:567-570; Kostrub C.F. et al. (1998) EMBO J. 17:2055-2066). These genes are involved in the mitotic checkpoint, and are induced by either DNA damage or blockage of replication. Induction of DNA damage or replication block leads to loss of function of the Hus1 gene and subsequent cell death. Human homologs have been identified for most  
30 of the rad genes, including ATM and ATR, the human homologs of rad3p. Mutations in the ATM gene are correlated with the severe congenital disease ataxia-telangiectasia (Savitsky, K. et al. (1995) Science 268:1749-1753). The human Hus1 protein has been shown to act in a complex with rad1 protein which interacts with rad9, making them central components of a DNA damage-responsive protein complex of human cells (Volkmer and Karnitz, *supra*).

35 The entry and exit of a cell from mitosis is regulated by the synthesis and destruction of a

family of activating proteins called cyclins. Cyclins act by binding to and activating a group of cyclin-dependent protein kinases (Cdks) which then phosphorylate and activate selected proteins involved in the mitotic process. Cyclins are characterized by a large region of shared homology that is approximately 180 amino acids in length and referred to as the "cyclin box" (Chapman, D.L. and D.J. Wolgemuth (1993) *Development* 118:229-240). In addition, cyclins contain a conserved 9 amino acid sequence in the N-terminal region of the molecule called the "destruction box." This sequence is believed to be a recognition code that triggers ubiquitin-mediated degradation of cyclin B (Hunt, T. (1991) *Nature* 349:100-101). Several types of cyclins exist (Ciechanover, A. (1994) *Cell* 79:13-21). Progression through G1 and S phase is driven by the G1 cyclins and their catalytic subunits, including Cdk2-cyclin A, Cdk2-cyclin E, Cdk4-cyclin D and Cdk6-cyclin D. Progression through the G2-M transition is driven by the activation of mitotic CDK-cyclin complexes such as Cdc2-cyclin A, Cdc2-cyclin B1 and Cdc2-cyclin B2 complexes (reviewed in Yang, J. and S. Kornbluth (1999) *Trends Cell Biol.* 9:207-210).

Cyclins are degraded through the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins in eukaryotic cells and in some bacteria. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that control cellular processes such as gene transcription and cell cycle progression. The UCS is implicated in the degradation of mitotic cyclin kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, *supra*).

The process of ubiquitin conjugation and protein degradation occurs in five principle steps (Jentsch, S. (1992) *Annu. Rev. Genet.* 26:179-207). First ubiquitin (Ub), a small, heat stable protein is activated by a ubiquitin-activating enzyme (E1) in an ATP dependent reaction which binds the C-terminus of Ub to the thiol group of an internal cysteine residue in E1. Second, activated Ub is transferred to one of several Ub-conjugating enzymes (E2). Different ubiquitin-dependent proteolytic pathways employ structurally similar, but distinct ubiquitin-conjugating enzymes that are associated with recognition subunits which direct them to proteins carrying a particular degradation signal. Third, E2 transfers the Ub molecule through its C-terminal glycine to a member of the ubiquitin-protein ligase family, E3. Fourth, E3 transfers the Ub molecule to the target protein. Additional Ub molecules may be added to the target protein forming a multi-Ub chain structure. Fifth, the ubiquitinated protein is then recognized and degraded by the proteasome, a large, multisubunit proteolytic enzyme complex, and Ub is released for re-utilization.

Prior to activation, Ub is usually expressed as a fusion protein composed of an N-terminal ubiquitin and a C-terminal extension protein (CEP) or as a polyubiquitin protein with Ub monomers attached head to tail. CEPs have characteristics of a variety of regulatory proteins; most are highly

basic, contain up to 30% lysine and arginine residues, and have nucleic acid-binding domains (Monia, B.P. et al. (1989) J. Biol. Chem. 264:4093-4103). The fusion protein is an important intermediate which appears to mediate co-regulation of the cell's translational and protein degradation activities, as well as localization of the inactive enzyme to specific cellular sites. Once  
5 delivered, C-terminal hydrolases cleave the fusion protein to release a functional Ub (Monia et al., *supra*).

Ub-conjugating enzymes (E2s) are important for substrate specificity in different UCS pathways. All E2s have a conserved domain of approximately 16 kDa called the UBC domain that is at least 35% identical in all E2s and contains a centrally located cysteine residue required for  
10 ubiquitin-enzyme thiolester formation (Jentsch, *supra*). A well conserved proline-rich element is located N-terminal to the active cysteine residue. Structural variations beyond this conserved domain are used to classify the E2 enzymes. Class I E2s consist almost exclusively of the conserved UBC domain. Class II E2s have various unrelated C-terminal extensions that contribute to substrate specificity and cellular localization. Class III E2s have unique N-terminal extensions which are  
15 believed to be involved in enzyme regulation or substrate specificity.

A mitotic cyclin-specific E2 (E2-C) is characterized by the conserved UBC domain, an N-terminal extension of 30 amino acids not found in other E2s, and a 7 amino acid unique sequence adjacent to this extension. These characteristics together with the high affinity of E2-C for cyclin identify it as a new class of E2 (Aristarkhov, A. et al. (1996) Proc. Natl. Acad. Sci. 93:4294-99).

20 Ubiquitin-protein ligases (E3s) catalyze the last step in the ubiquitin conjugation process, covalent attachment of ubiquitin to the substrate. E3 plays a key role in determining the specificity of the process. Only a few E3s have been identified so far. One type of E3 ligases is the HECT (homologous to E6-AP C-terminus) domain protein family. One member of the family, E6-AP (E6-associated protein) is required, along with the human papillomavirus (HPV) E6 oncoprotein, for  
25 the ubiquitination and degradation of p53 (Scheffner, M. et al. (1993) Cell 75:495-505). The C-terminal domain of HECT proteins contains the highly conserved ubiquitin-binding cysteine residue. The N-terminal region of the various HECT proteins is variable and is believed to be involved in specific substrate recognition (Huibregtse, J.M. et al. (1997) Proc. Natl Acad. Sci. USA 94:3656-3661). The SCF (Skp1-Cdc53/Cullin-F box receptor) family of proteins comprise another  
30 group of ubiquitin ligases (Deshaies, R. (1999) Annu. Rev. Dev. Biol. 15:435-467). Multiple proteins are recruited into the SCF complex, including Skp1, cullin, and an F box domain containing protein. The F box protein binds the substrate for the ubiquitination reaction and may play roles in determining substrate specificity and orienting the substrate for reaction. Skp1 interacts with both the F box protein and cullin and may be involved in positioning the F box protein and cullin in the  
35 complex for transfer of ubiquitin from the E2 enzyme to the protein substrate. Substrates of SCF

ligases include proteins involved in regulation of CDK activity, activation of transcription, signal transduction, assembly of kinetochores, and DNA replication.

Sgt1 was identified in a screen for genes in yeast that suppress defects in kinetochore function caused by mutations in Skp1 (Kitagawa, K. et al. (1999) Mol. Cell 4:21-33). Sgt1 interacts  
5 with Skp1 and associates with SCF ubiquitin ligase. Defects in Sgt1 cause arrest of cells at either G1 or G2 stages of the cell cycle. A yeast Sgt1 null mutant can be rescued by human Sgt1, an indication of the conservation of Sgt1 function across species. Sgt1 is required for assembly of kinetochore complexes in yeast.

Abnormal activities of the UCS are implicated in a number of diseases and disorders. These  
10 include, e.g., cachexia (Llovera, M. et al. (1995) Int. J. Cancer 61:138-141), degradation of the tumor-suppressor protein, p53 (Ciechanover, *supra*), and neurodegeneration such as observed in Alzheimer's disease (Gregori, L. et al. (1994) Biochem. Biophys. Res. Commun. 203:1731-1738). Since ubiquitin conjugation is a rate-limiting step in antigen presentation, the ubiquitin degradation pathway may also have a critical role in the immune response (Grant, E.P. et al. (1995) J. Immunol.  
15 155:3750-3758).

Certain cell proliferation disorders can be identified by changes in the protein complexes that normally control progression through the cell cycle. A primary treatment strategy involves reestablishing control over cell cycle progression by manipulation of the proteins involved in cell cycle regulation (Nigg, E.A. (1995) BioEssays 17:471-480).

## 20 **Embryogenesis**

Mammalian embryogenesis is a process which encompasses the first few weeks of development following conception. During this period, embryogenesis proceeds from a single fertilized egg to the formation of the three embryonic tissues, then to an embryo which has most of its internal organs and all of its external features.

25 The normal course of mammalian embryogenesis depends on the correct temporal and spatial regulation of a large number of genes and tissues. These regulation processes have been intensely studied in mouse. An essential process that is still poorly understood is the activation of the embryonic genome after fertilization. As mouse oocytes grow, they accumulate transcripts that are either translated directly into proteins or stored for later activation by regulated polyadenylation.  
30 During subsequent meiotic maturation and ovulation, the maternal genome is transcriptionally inert, and most maternal transcripts are deadenylated and/or degraded prior to, or together with, the activation of the zygotic genes at the two-cell stage (Stutz, A. et al. (1998) Genes Dev. 12:2535-2548). The maternal to embryonic transition involves the degradation of oocyte, but not zygotic transcripts, the activation of the embryonic genome, and the induction of cell cycle progression to  
35 accommodate early development.



MATER (Maternal Antigen That Embryos Require) was initially identified as a target of antibodies from mice with ovarian immunity (Tong, Z-B. and L.M. Nelson (1999) Endocrinology 140:3720-3726). Expression of the gene encoding MATER is restricted to the oocyte, making it one of a limited number of known maternal-effect genes in mammals (Tong, Z-B. et al. (2000) Mamm. 5 Genome 11:281-287). The MATER protein is required for embryonic development beyond two cells, based upon preliminary results from mice in which this gene has been inactivated. The 1111-amino acid MATER protein contains a hydrophilic repeat region in the amino terminus, and a region containing 14 leucine-rich repeats in the carboxyl terminus. These repeats resemble the sequence found in porcine ribonuclease inhibitor that is critical for protein-protein interactions.

10 The degradation of maternal transcripts during meiotic maturation and ovulation may involve the activation of a ribonuclease just prior to ovulation. Thus the function of MATER may be to bind to the maternal ribonuclease and prevent degradation of zygotic transcripts (Tong et al., *supra*). In addition to its role in oocyte development and embryogenesis, MATER may also be relevant to the pathogenesis of ovarian immunity, as it is a target of autoantibodies in mice with autoimmune 15 oophoritis (Tong and Nelson, *supra*).

The maternal mRNA D7 is a moderately abundant transcript in *Xenopus laevis* whose expression is highest in, and perhaps restricted to, oogenesis and early embryogenesis. The D7 protein is absent from oocytes and first begins to accumulate during oocyte maturation. Its levels are highest during the first day of embryonic development and then they decrease. The loss of D7 20 protein affects the maturation process itself, significantly delaying the time course of germinal vesicle breakdown. Thus, D7 is a newly described protein involved in oocyte maturation (Smith, R.C. et al. (1988) Genes Dev. 2(10):1296-306.)

Many other genes are involved in subsequent stages of embryogenesis. After fertilization, the oocyte is guided by fimbria at the distal end of each fallopian tube into and through the fallopian tube 25 and thence into the uterus. Changes in the uterine endometrium prepare the tissue to support the implantation and embryonic development of a fertilized ovum. Several stages of division have occurred before the dividing ovum, now a blastocyst with about 100 cells, enters the uterus. Upon reaching the uterus, the developing blastocyst usually remains in the uterine cavity an additional two to four days before implanting in the endometrium, the inner lining of the uterus. Implantation results 30 from the action of trophoblast cells that develop over the surface of the blastocyst. These cells secrete proteolytic enzymes that digest and liquefy the cells of the endometrium. The invasive process is reviewed in Fisher, S.J. and C.H. Damsky (1993; Semin Cell Biol 4:183-188) and Graham, C.H. and P.K. Lala (1992; Biochem Cell Biol 70:867-874). Once implantation has taken place, the trophoblast and other sublying cells proliferate rapidly, forming the placenta and the various 35 membranes of pregnancy. (See Guyton, A.C. (1991) Textbook of Medical Physiology, 8<sup>th</sup> ed. W.B.

Saunders Company, Philadelphia PA, pp. 915-919.)

The placenta has an essential role in protecting and nourishing the developing fetus. In most species the syncytiotrophoblast layer is present on the outside of the placenta at the fetal-maternal interface. This is a continuous structure, one cell deep, formed by the fusion of the constituent  
5 trophoblast cells. The syncytiotrophoblast cells play important roles in maternal-fetal exchange, in tissue remodeling during fetal development, and in protecting the developing fetus from the maternal immune response (Stoye, J.P. and J.M. Coffin (2000) Nature 403:715-717).

A gene called syncytin is the envelope gene of a human endogenous defective provirus. Syncytin is expressed in high levels in placenta, and more weakly in testis, but is not detected in any  
10 other tissues (Mi, S. et al. (2000) Nature 403:785-789). Syncytin expression in the placenta is restricted to the syncytiotrophoblasts. Since retroviral *env* proteins are often involved in promoting cell fusion events, it was thought that syncytin might be involved in regulating the fusion of trophoblast cells into the syncytiotrophoblast layer. Experiments demonstrated that syncytin can mediate cell fusion *in vitro*, and that anti-syncytin antibodies can inhibit the fusion of placental  
15 cytotrophoblasts (Mi et al., *supra*). In addition, a conserved immunosuppressive domain present in retroviral envelope proteins, and found in syncytin at amino acid residues 373-397, might be involved in preventing maternal immune responses against the developing embryo.

Syncytin may also be involved in regulating trophoblast invasiveness by inducing trophoblast fusion and terminal differentiation (Mi et al., *supra*). Insufficient trophoblast infiltration of the  
20 uterine wall is associated with placental disorders such as preeclampsia, or pregnancy induced hypertension, while uncontrolled trophoblast invasion is observed in choriocarcinoma and other gestational trophoblastic diseases. Thus syncytin function may be involved in these diseases.

### Cell Differentiation

Multicellular organisms are comprised of diverse cell types that differ dramatically both in  
25 structure and function, despite the fact that each cell is like the others in its hereditary endowment. Cell differentiation is the process by which cells come to differ in their structure and physiological function. The cells of a multicellular organism all arise from mitotic divisions of a single-celled zygote. The zygote is totipotent, meaning that it has the ability to give rise to every type of cell in the adult body. During development the cellular descendants of the zygote lose their totipotency and  
30 become determined. Once its prospective fate is achieved, a cell is said to have differentiated. All descendants of this cell will be of the same type.

Human growth and development requires the spatial and temporal regulation of cell differentiation, along with cell proliferation and regulated cell death. These processes coordinate to control reproduction, aging, embryogenesis, morphogenesis, organogenesis, and tissue repair and  
35 maintenance. The processes involved in cell differentiation are also relevant to disease states such as

cancer, in which case the factors regulating normal cell differentiation have been altered, allowing the cancerous cells to proliferate in an anaplastic, or undifferentiated, state.

The mechanisms of differentiation involve cell-specific regulation of transcription and translation, so that different genes are selectively expressed at different times in different cells.

5 Genetic experiments using the fruit fly *Drosophila melanogaster* have identified regulated cascades of transcription factors which control pattern formation during development and differentiation. These include the homeotic genes, which encode transcription factors containing homeobox motifs. The products of homeotic genes determine how the insect's imaginal discs develop from masses of undifferentiated cells to specific segments containing complex organs. Many genes found to be  
10 involved in cell differentiation and development in *Drosophila* have homologs in mammals. Some human genes have equivalent developmental roles to their *Drosophila* homologs. The human homolog of the *Drosophila* eyes absent gene (*eya*) underlies branchio-oto-renal syndrome, a developmental disorder affecting the ears and kidneys (Abdelhak, S. et al. (1997) Nat. Genet. 15:157-164). The *Drosophila* slit gene encodes a secreted leucine-rich repeat containing protein expressed  
15 by the midline glial cells and required for normal neural development.

At the cellular level, growth and development are governed by the cell's decision to enter into or exit from the cell cycle and by the cell's commitment to a terminally differentiated state. Differential gene expression within cells is triggered in response to extracellular signals and other environmental cues. Such signals include growth factors and other mitogens such as retinoic acid;  
20 cell-cell and cell-matrix contacts; and environmental factors such as nutritional signals, toxic substances, and heat shock. Candidate genes that may play a role in differentiation can be identified by altered expression patterns upon induction of cell differentiation *in vitro*.

The final step in cell differentiation results in a specialization that is characterized by the production of particular proteins, such as contractile proteins in muscle cells, serum proteins in liver  
25 cells and globins in red blood cell precursors. The expression of these specialized proteins depends at least in part on cell-specific transcription factors. For example, the homeobox-containing transcription factor PAX-6 is essential for early eye determination, specification of ocular tissues, and normal eye development in vertebrates.

In the case of epidermal differentiation, the induction of differentiation-specific genes occurs  
30 either together with or following growth arrest and is believed to be linked to the molecular events that control irreversible growth arrest. Irreversible growth arrest is an early event which occurs when cells transit from the basal to the innermost suprabasal layer of the skin and begin expressing squamous-specific genes. These genes include those involved in the formation of the cross-linked envelope, such as transglutaminase I and III, involucrin, loricin, and small proline-rich repeat (SPRR)  
35 proteins. The SPRR proteins are 8-10 kDa in molecular mass, rich in proline, glutamine, and

cysteine, and contain similar repeating sequence elements. The SPRR proteins may be structural proteins with a strong secondary structure or metal-binding proteins such as metallothioneins. (Jetten, A.M. and B.L. Harvat (1997) J. Dermatol. 24:711-725; PRINTS Entry PR00021 PRORICH Small proline-rich protein signature.)

5       The Wnt gene family of secreted signaling molecules is highly conserved throughout eukaryotic cells. Members of the Wnt family are involved in regulating chondrocyte differentiation within the cartilage template. Wnt-5a, Wnt-5b and Wnt-4 genes are expressed in chondrogenic regions of the chicken limb, Wnt-5a being expressed in the perichondrium (mesenchymal cells immediately surrounding the early cartilage template). Wnt-5a misexpression delays the maturation  
10 of chondrocytes and the onset of bone collar formation in chicken limb (Hartmann, C. and C.J. Tabin (2000) Development 127:3141-3159).

Glypicans are a family of cell surface heparan sulfate proteoglycans that play an important role in cellular growth control and differentiation. Cerebroglycan, a heparan sulfate proteoglycan expressed in the nervous system, is involved with the motile behavior of developing neurons (Stipp,  
15 C.S. et al. (1994) J. Cell Biol. 124:149-160).

Notch plays an active role in the differentiation of glial cells, and influences the length and organization of neuronal processes (for a review, see Frisen, J. and U. Lendahl (2001) Bioessays 23:3-7). The Notch receptor signaling pathway is important for morphogenesis and development of many organs and tissues in multicellular species. *Drosophila* fringe proteins modulate the activation  
20 of the Notch signal transduction pathway at the dorsal-ventral boundary of the wing imaginal disc. Mammalian fringe-related family members participate in boundary determination during segmentation (Johnston, S.H. et al. (1997) Development 124:2245-2254).

Recently a number of proteins have been found to contain a conserved cysteine-rich domain of about 60 amino-acid residues called the LIM domain (for Lin-11 Isl-1 Mec-3) (Freyd, G. et al.  
25 (1990) Nature 344:876-879; Baltz, R. et al. (1992) Plant Cell 4:1465-1466). In the LIM domain, there are seven conserved cysteine residues and a histidine. The LIM domain binds two zinc ions (Michelsen, J.W. et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90:4404-4408). LIM does not bind DNA; rather, it seems to act as an interface for protein-protein interaction.

### Apoptosis

30       Apoptosis is the genetically controlled process by which unneeded or defective cells undergo programmed cell death. Selective elimination of cells is as important for morphogenesis and tissue remodeling as is cell proliferation and differentiation. Lack of apoptosis may result in hyperplasia and other disorders associated with increased cell proliferation. Apoptosis is also a critical component of the immune response. Immune cells such as cytotoxic T-cells and natural killer cells  
35 prevent the spread of disease by inducing apoptosis in tumor cells and virus-infected cells. In

addition, immune cells that fail to distinguish self molecules from foreign molecules must be eliminated by apoptosis to avoid an autoimmune response.

Apoptotic cells undergo distinct morphological changes. Hallmarks of apoptosis include cell shrinkage, nuclear and cytoplasmic condensation, and alterations in plasma membrane topology.

- 5 Biochemically, apoptotic cells are characterized by increased intracellular calcium concentration, fragmentation of chromosomal DNA, and expression of novel cell surface components.

The molecular mechanisms of apoptosis are highly conserved, and many of the key protein regulators and effectors of apoptosis have been identified. Apoptosis generally proceeds in response to a signal which is transduced intracellularly and results in altered patterns of gene expression and  
10 protein activity. Signaling molecules such as hormones and cytokines are known both to stimulate and to inhibit apoptosis through interactions with cell surface receptors. Transcription factors also play an important role in the onset of apoptosis. A number of downstream effector molecules, especially proteases, have been implicated in the degradation of cellular components and the proteolytic activation of other apoptotic effectors.

- 15 The Bcl-2 family of proteins, as well as other cytoplasmic proteins, are key regulators of apoptosis. There are at least 15 Bcl-2 family members within 3 subfamilies. These proteins have been identified in mammalian cells and in viruses, and each possesses at least one of four Bcl-2 homology domains (BH1 to BH4), which are highly conserved. Bcl-2 family proteins contain the BH1 and BH2 domains, which are found in members of the pro-survival subfamily, while those  
20 proteins which are most similar to Bcl-2 have all four conserved domains, enabling inhibition of apoptosis following encounters with a variety of cytotoxic challenges. Members of the pro-survival subfamily include Bcl-2, Bcl-x<sub>L</sub>, Bcl-w, Mcl-1, and A1 in mammals; NF-13 (chicken); CED-9 (*Caenorhabditis elegans*); and viral proteins BHRF1, LMW5-HL, ORF16, KS-Bcl-2, and E1B-19K. The BH3 domain is essential for the function of pro-apoptosis subfamily proteins. The two pro-  
25 apoptosis subfamilies, Bax and BH3, include Bax, Bak, and Bok (also called Mtd); and Bik, Blk, Hrk, BNIP3, Bim<sub>L</sub>, Bad, Bid, and Egl-1 (*C. elegans*); respectively. Members of the Bax subfamily contain the BH1, BH2, and BH3 domains, and resemble Bcl-2 rather closely. In contrast, members of the BH3 subfamily have only the 9-16 residue BH3 domain, being otherwise unrelated to any known protein, and only Bik and Blk share sequence similarity. The proteins of the two pro-apoptosis  
30 subfamilies may be the antagonists of pro-survival subfamily proteins. This is illustrated in *C. elegans* where Egl-1, which is required for apoptosis, binds to and acts via CED-9 (for review, see Adams, J.M. and S. Cory (1998) Science 281:1322-1326).

- Heterodimerization between pro-apoptosis and anti-apoptosis subfamily proteins seems to have a titrating effect on the functions of these protein subfamilies, which suggests that relative  
35 concentrations of the members of each subfamily may act to regulate apoptosis. Heterodimerization

is not required for a pro-survival protein; however, it is essential in the BH3 subfamily, and less so in the Bax subfamily.

The Bcl-2 protein has 2 isoforms, alpha and beta, which are formed by alternative splicing. It forms homodimers and heterodimers with Bax and Bak proteins and the Bcl-X isoform Bcl-x<sub>S</sub>.

- 5 Heterodimerization with Bax requires intact BH1 and BH2 domains, and is necessary for pro-survival activity. The BH4 domain seems to be involved in pro-survival activity as well. Bcl-2 is located within the inner and outer mitochondrial membranes, as well as within the nuclear envelope and endoplasmic reticulum, and is expressed in a variety of tissues. Its involvement in follicular lymphoma (type II chronic lymphatic leukemia) is seen in a chromosomal translocation T(14;18)  
10 (q32;q21) and involves immunoglobulin gene regions.

The Bcl-x protein is a dominant regulator of apoptotic cell death. Alternative splicing results in three isoforms, Bcl-xB, a long isoform, and a short isoform. The long isoform exhibits cell death repressor activity, while the short isoform promotes apoptosis. Bcl-xL forms heterodimers with Bax and Bak, although heterodimerization with Bax does not seem to be necessary for pro-survival (anti-  
15 apoptosis) activity. Bcl-xS forms heterodimers with Bcl-2. Bcl-x is found in mitochondrial membranes and the perinuclear envelope. Bcl-xS is expressed at high levels in developing lymphocytes and other cells undergoing a high rate of turnover. Bcl-xL is found in adult brain and in other tissues' long-lived post-mitotic cells. As with Bcl-2, the BH1, BH2, and BH4 domains are involved in pro-survival activity.

- 20 The Bcl-w protein is found within the cytoplasm of almost all myeloid cell lines and in numerous tissues, with the highest levels of expression in brain, colon, and salivary gland. This protein is expressed in low levels in testis, liver, heart, stomach, skeletal muscle, and placenta, and a few lymphoid cell lines. Bcl-w contains the BH1, BH2, and BH4 domains, all of which are needed for its cell survival promotion activity. Although mice in which Bcl-w gene function was disrupted  
25 by homologous recombination were viable, healthy, and normal in appearance, and adult females had normal reproductive function, the adult males were infertile. In these males, the initial, prepuberty stage of spermatogenesis was largely unaffected and the testes developed normally. However, the seminiferous tubules were disorganized, contained numerous apoptotic cells, and were incapable of producing mature sperm. This mouse model may be applicable to some cases of human male sterility  
30 and suggests that alteration of programmed cell death in the testes may be useful in modulating fertility (Print, C.G. et al. (1998) Proc. Natl. Acad. Sci. USA 95:12424-12431).

- Studies in rat ischemic brain found Bcl-w to be overexpressed relative to its normal low constitutive level of expression in nonischemic brain. Furthermore, *in vitro* studies to examine the mechanism of action of Bcl-w revealed that isolated rat brain mitochondria were unable to respond to  
35 an addition of recombinant Bax or high concentrations of calcium when Bcl-w was also present. The

normal response would be the release of cytochrome c from the mitochondria. Additionally, recombinant Bcl-w protein was found to inhibit calcium-induced loss of mitochondrial transmembrane potential, which is indicative of permeability transition. Together these findings suggest that Bcl-w may be a neuro-protectant against ischemic neuronal death and may achieve this protection via the mitochondrial death-regulatory pathway (Yan, C. et al. (2000) *J. Cereb. Blood Flow Metab.* 20:620-630).

The bfl-1 gene is an additional member of the Bcl-2 family, and is also a suppressor of apoptosis. The Bfl-1 protein has 175 amino acids, and contains the BH1, BH2, and BH3 conserved domains found in Bcl-2 family members. It also contains a Gln-rich NH2-terminal region and lacks an NH domain 1, unlike other Bcl-2 family members. The mouse A1 protein shares high sequence homology with Bfl-1 and has the 3 conserved domains found in Bfl-1. Apoptosis induced by the p53 tumor suppressor protein is suppressed by Bfl-1, similar to the action of Bcl-2, Bcl-xL, and EBV-BHRF1 (D'Sa-Eipper, C. et al. (1996) *Cancer Res.* 56:3879-3882). Bfl-1 is found intracellularly, with the highest expression in the hematopoietic compartment, i.e. blood, spleen, and bone marrow; moderate expression in lung, small intestine, and testis; and minimal expression in other tissues. It is also found in vascular smooth muscle cells and hematopoietic malignancies. A correlation has been noted between the expression level of bfl-1 and the development of stomach cancer, suggesting that the Bfl-1 protein is involved in the development of stomach cancer, either in the promotion of cancerous cell survival or in cancer (Choi, S.S. et al. (1995) *Oncogene* 11:1693-1698).

Cancers are characterized by continuous or uncontrolled cell proliferation. Some cancers are associated with suppression of normal apoptotic cell death. Strategies for treatment may involve either reestablishing control over cell cycle progression, or selectively stimulating apoptosis in cancerous cells (Nigg, E.A. (1995) *BioEssays* 17:471-480). Immunological defenses against cancer include induction of apoptosis in mutant cells by tumor suppressors, and the recognition of tumor antigens by T lymphocytes. Response to mitogenic stresses is frequently controlled at the level of transcription and is coordinated by various transcription factors. For example, the Rel/NF-kappa B family of vertebrate transcription factors plays a pivotal role in inflammatory and immune responses to radiation. The NF-kappa B family includes p50, p52, RelA, RelB, cRel, and other DNA-binding proteins. The p52 protein induces apoptosis, upregulates the transcription factor c-Jun, and activates c-Jun N-terminal kinase 1 (JNK1) (Sun, L. et al. (1998) *Gene* 208:157-166). Most NF-kappa B proteins form DNA-binding homodimers or heterodimers. Dimerization of many transcription factors is mediated by a conserved sequence known as the bZIP domain, characterized by a basic region followed by a leucine zipper.

The Fas/Apo-1 receptor (FAS) is a member of the tumor necrosis factor (TNF) receptor family. Upon binding its ligand (Fas ligand), the membrane-spanning FAS induces apoptosis by

recruiting several cytoplasmic proteins that transmit the death signal. One such protein, termed FAS-associated protein factor 1 (FAF1), was isolated from mice, and it was demonstrated that expression of FAF1 in L cells potentiated FAS-induced apoptosis (Chu, K. et al. (1995) Proc. Natl. Acad. Sci. USA 92:11894-11898). Subsequently, FAS-associated factors have been isolated from numerous  
5 other species, including fruit fly and quail (Frohlich, T. et al. (1998) J. Cell Sci. 111:2353-2363). Another cytoplasmic protein that functions in the transmittal of the death signal from Fas is the Fas-associated death domain protein, also known as FADD. FADD transmits the death signal in both FAS-mediated and TNF receptor-mediated apoptotic pathways by activating caspase-8 (Bang, S. et al. (2000) J. Biol. Chem. 275:36217-36222).

10 Fragmentation of chromosomal DNA is one of the hallmarks of apoptosis. DNA fragmentation factor (DFF) is a protein composed of two subunits, a 40-kDa caspase-activated nuclease termed DFF40/CAD, and its 45-kDa inhibitor DFF45/ICAD. Two mouse homologs of DFF45/ICAD, termed CIDE-A and CIDE-B, have recently been described (Inohara, N. et al. (1998) EMBO J. 17:2526-2533). CIDE-A and CIDE-B expression in mammalian cells activated apoptosis,  
15 while expression of CIDE-A alone induced DNA fragmentation. In addition, FAS-mediated apoptosis was enhanced by CIDE-A and CIDE-B, further implicating these proteins as effectors that mediate apoptosis.

Transcription factors play an important role in the onset of apoptosis. A number of downstream effector molecules, particularly proteases such as the cysteine proteases called caspases,  
20 are involved in the initiation and execution phases of apoptosis. The activation of the caspases results from the competitive action of the pro-survival and pro-apoptosis Bcl-2-related proteins (Print, C.G. et al. (1998) Proc. Natl. Acad. Sci. USA 95:12424-12431). A pro-apoptotic signal can activate initiator caspases that trigger a proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death of the cell. Two active site residues, a cysteine and a  
25 histidine, have been implicated in the catalytic mechanism. Caspases are among the most specific endopeptidases, cleaving after aspartate residues.

Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain interacts with cofactors that can positively or negatively affect apoptosis. An activating  
30 signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1 numbering convention) and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long prodomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. Some caspases contain a "death effector domain" in their prodomain  
35 by which they can be recruited into self-activating complexes with other caspases and FADD protein-



associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer.

Tumor necrosis factor (TNF) and related cytokines induce apoptosis in lymphoid cells. (Reviewed in Nagata, S. (1997) *Cell* 88:355-365.) Binding of TNF to its receptor triggers a signal  
5 transduction pathway that results in the activation of a proteolytic caspase cascade. One such caspase, ICE (Interleukin-1 $\beta$  converting enzyme), is a cysteine protease comprised of two large and two small subunits generated by ICE auto-cleavage (Dinarello, C.A. (1994) *FASEB J.* 8:1314-1325). ICE is expressed primarily in monocytes. ICE processes the cytokine precursor, interleukin-1 $\beta$ , into its active form, which plays a central role in acute and chronic inflammation, bone resorption,  
10 myelogenous leukemia, and other pathological processes. ICE and related caspases cause apoptosis when overexpressed in transfected cell lines.

A caspase recruitment domain (CARD) is found within the prodomain of several apical caspases and is conserved in several apoptosis regulatory molecules such as Apaf-2, RAIDD, and cellular inhibitors of apoptosis proteins (IAPs) (Hofmann, K. et al. (1997) *Trends Biochem. Sci.*  
15 22:155-157). The regulatory role of CARD in apoptosis may be to allow proteins such as Apaf-1 to associate with caspase-9 (Li, P. et al. (1997) *Cell* 91:479-489). A human cDNA encoding an apoptosis repressor with a CARD (ARC) which is expressed in both skeletal and cardiac muscle has been identified and characterized. ARC functions as an inhibitor of apoptosis and interacts selectively with caspases (Koseki, T. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:5156-5160). All of  
20 these interactions have clear effects on the control of apoptosis (reviewed in Chan S.L. and M.P. Mattson (1999) *J. Neurosci. Res.* 58:167-190; Salveson, G.S. and V.M. Dixit (1999) *Proc. Natl. Acad. Sci. USA* 96:10964-10967).

ES18 was identified as a potential regulator of apoptosis in mouse T-cells (Park, E.J. et al. (1999) *Nuc. Acid. Res.* 27:1524-1530). ES18 is 428 amino acids in length, contains an N-terminal  
25 proline-rich region, an acidic glutamic acid-rich domain, and a putative LXXLL nuclear receptor binding motif. The protein is preferentially expressed in lymph nodes and thymus. The level of ES18 expression increases in T-cell thymoma S49.1 in response to treatment with dexamethasone, staurosporine, or C2-ceramide, which induce apoptosis. ES18 may play a role in stimulating apoptotic cell death in T-cells.

30 The rat ventral prostate (RVP) is a model system for the study of hormone-regulated apoptosis. RVP epithelial cells undergo apoptosis in response to androgen deprivation. Messenger RNA (mRNA) transcripts that are up-regulated in the apoptotic RVP have been identified (Briehl, M. M. and R.L. Miesfeld (1991) *Mol. Endocrinol.* 5:1381-1388). One such transcript encodes RVP.1, the precise role of which in apoptosis has not been determined. The human homolog of RVP.1,  
35 hRVP1, is 89% identical to the rat protein (Katahira, J. et al. (1997) *J. Biol. Chem.* 272:26652-

26658). hRVP1 is 220 amino acids in length and contains four transmembrane domains. hRVP1 is highly expressed in the lung, intestine, and liver. Interestingly, hRVP1 functions as a low affinity receptor for the *Clostridium perfringens* enterotoxin, a causative agent of diarrhea in humans and other animals.

5 Cytokine-mediated apoptosis plays an important role in hematopoiesis and the immune response. Myeloid cells, which are the stem cell progenitors of macrophages, neutrophils, erythrocytes, and other blood cells, proliferate in response to specific cytokines such as granulocyte/macrophage-colony stimulating factor (GM-CSF) and interleukin-3 (IL-3). When deprived of GM-CSF or IL-3, myeloid cells undergo apoptosis. The murine *requiem* (*req*) gene  
10 encodes a putative transcription factor required for this apoptotic response in the myeloid cell line FDCP-1 (Gabig, T. G. et al. (1994) J. Biol. Chem. 269:29515-29519). The Req protein is 371 amino acids in length and contains a nuclear localization signal, a single *Kruppel*-type zinc finger, an acidic domain, and a cluster of four unique zinc-finger motifs enriched in cysteine and histidine residues involved in metal binding. Expression of *req* is not myeloid- or apoptosis-specific, suggesting that  
15 additional factors regulate Req activity in myeloid cell apoptosis.

Dysregulation of apoptosis has recently been recognized as a significant factor in the pathogenesis of many human diseases. For example, excessive cell survival caused by decreased apoptosis can contribute to disorders related to cell proliferation and the immune response. Such disorders include cancer, autoimmune diseases, viral infections, and inflammation. In contrast,  
20 excessive cell death caused by increased apoptosis can lead to degenerative and immunodeficiency disorders such as AIDS, neurodegenerative diseases, and myelodysplastic syndromes. (Thompson, C.B. (1995) Science 267:1456-1462.)

Impaired regulation of apoptosis is also associated with loss of neurons in Alzheimer's disease. Alzheimer's disease is a progressive neurodegenerative disorder that is characterized by the  
25 formation of senile plaques and neurofibrillary tangles containing amyloid beta peptide. These plaques are found in limbic and association cortices of the brain, including hippocampus, temporal cortices, cingulate cortex, amygdala, nucleus basalis and locus caeruleus. B-amyloid peptide participates in signaling pathways that induce apoptosis and lead to the death of neurons (Kajkowski, C. et al. (2001) J. Biol. Chem. 276:18748-18756). Early in Alzheimer's pathology, physiological  
30 changes are visible in the cingulate cortex (Minoshima, S. et al. (1997) Annals of Neurology 42:85-94). In subjects with advanced Alzheimer's disease, accumulating plaques damage the neuronal architecture in limbic areas and eventually cripple the memory process.

#### Aging and Senescence

Studies of the aging process or senescence have shown a number of characteristic cellular  
35 and molecular changes (Fauci et al. (1998) Harrison's Principles of Internal Medicine, McGraw-Hill,

New York NY, p.37). These characteristics include increases in chromosome structural abnormalities, DNA cross-linking, incidence of single-stranded breaks in DNA, losses in DNA methylation, and degradation of telomere regions. In addition to these DNA changes, post-translational alterations of proteins increase including, deamidation, oxidation, cross-linking, and nonenzymatic glycation. Still further molecular changes occur in the mitochondria of aging cells through deterioration of structure. These changes eventually contribute to decreased function in every organ of the body.

### **Cancer**

Understanding of the neoplastic process can be aided by the identification of molecular markers of prognostic and diagnostic importance. Cancers are associated with oncoproteins which are capable of transforming normal cells into malignant cells. Some oncoproteins are mutant isoforms of the normal protein while others are abnormally expressed with respect to location or level of expression. Normal cell proliferation begins with binding of a growth factor to its receptor on the cell membrane, resulting in activation of a signal system that induces and activates nuclear regulatory factors to initiate DNA transcription, subsequently leading to cell division. Classes of oncoproteins known to affect the cell cycle controls include growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. Several types of cancer-specific genetic markers, such as tumor antigens and tumor suppressors, have also been identified.

### **Oncogenes**

Oncoproteins are encoded by genes, called oncogenes, that are derived from genes that normally control cell growth and development. Many oncogenes have been identified and characterized. These include growth factors such as *sis*, receptors such as *erbA*, *erbB*, *neu*, and *ros*, intracellular receptors such as *src*, *yes*, *fps*, *abl*, and *met*, protein-serine/threonine kinases such as *mos* and *raf*, nuclear transcription factors such as *jun*, *fos*, *myc*, *N-myc*, *myb*, *ski*, and *rel*, cell cycle control proteins such as *RB* and *p53*, mutated tumor-suppressor genes such as *mdm2*, *Cip1*, *p16*, and *cyclin D*, *ras*, *set*, *can*, *sec*, and *gag R10*.

Viral oncogenes are integrated into the human genome after infection of human cells by certain viruses. Examples of viral oncogenes include v-src, v-abl, and v-fps. Transformation of normal genes to oncogenes may also occur by chromosomal translocation. The Philadelphia chromosome, characteristic of chronic myeloid leukemia and a subset of acute lymphoblastic leukemias, results from a reciprocal translocation between chromosomes 9 and 22 that moves a truncated portion of the proto-oncogene *c-abl* to the breakpoint cluster region (bcr) on chromosome 22. The hybrid *c-abl-bcr* gene encodes a chimeric protein that has tyrosine kinase activity. In chronic myeloid leukemia, the chimeric protein has a molecular weight of 210 kd, whereas in acute leukemias

a more active 180 kd tyrosine kinase is formed (Robbins, S.L. et al. (1994) Pathologic Basis of Disease, W.B. Saunders Co., Philadelphia PA).

The Ras superfamily of small GTPases is involved in the regulation of a wide range of cellular signaling pathways. Ras family proteins are membrane-associated proteins acting as molecular switches that bind GTP and GDP, hydrolyzing GTP to GDP. In the active GTP-bound state Ras family proteins interact with a variety of cellular targets to activate downstream signaling pathways. For example, members of the Ras subfamily are essential in transducing signals from receptor tyrosine kinases (RTKs) to a series of serine/threonine kinases which control cell growth and differentiation. Activated Ras genes were initially found in human cancers and subsequent studies confirmed that Ras function is critical in the determination of whether cells continue to grow or become terminally differentiated (Barbacid, M. (1987) *Annu. Rev. Biochem.* 56:779-827; Treisman, R. (1994) *Curr. Opin. Genet. Dev.* 4:96-98). Mutant Ras proteins, which bind but can not hydrolyze GTP, are permanently activated, and cause continuous cell proliferation or cancer.

Activation of Ras family proteins is catalyzed by guanine nucleotide exchange factors (GEFs) which catalyze the dissociation of bound GDP and subsequent binding of GTP. A recently discovered RalGEF-like protein, RGL3, interacts with both Ras and the related protein Rit. Constitutively active Rit, like Ras, can induce oncogenic transformation, although since Rit fails to interact with most known Ras effector proteins, novel cellular targets may be involved in Rit transforming activity. RGL3 interacts with both Ras and Rit, and thus may act as a downstream effector for these proteins (Shao, H. and D.A. Andres (2000) *J. Biol. Chem.* 275:26914-26924).

#### Tumor antigens

Tumor antigens are cell surface molecules that are differentially expressed in tumor cells relative to non-tumor tissues. Tumor antigens make tumor cells immunologically distinct from normal cells and are potential diagnostics for human cancers. Several monoclonal antibodies have been identified which react specifically with cancerous cells such as T-cell acute lymphoblastic leukemia and neuroblastoma (Minegishi, M. et al. (1989) *Leukemia Res.* 13:43-51; Takagi, S. et al. (1995) *Int. J. Cancer* 61:706-715). In addition, the discovery of high level expression of the HER2 gene in breast tumors has led to the development of therapeutic treatments (Liu, E. et al. (1992) *Oncogene* 7: 1027-1032; Kern, J.A. (1993) *Am. J. Respir. Cell Mol. Biol.* 9:448-454). Tumor antigens are found on the cell surface and have been characterized either as membrane proteins or glycoproteins. For example, MAGE genes encode a family of tumor antigens recognized on melanoma cell surfaces by autologous cytolytic T lymphocytes. Among the 12 human MAGE genes isolated, half are differentially expressed in tumors of various histological types (De Plaen, E. et al. (1994) *Immunogenetics* 40:360-369). None of the 12 MAGE genes, however, is expressed in healthy tissues except testis and placenta.

### Tumor suppressors

Tumor suppressor genes are generally defined as genetic elements whose loss or inactivation contributes to the deregulation of cell proliferation and the pathogenesis and progression of cancer. Tumor suppressor genes normally function to control or inhibit cell growth in response to stress and  
5 to limit the proliferative life span of the cell. Several tumor suppressor genes have been identified including the genes encoding the retinoblastoma (Rb) protein, p53, and the breast cancer 1 and 2 proteins (BRCA1 and BRCA2). Mutations in these genes are associated with acquired and inherited genetic predisposition to the development of certain cancers.

The role of p53 in the pathogenesis of cancer has been extensively studied. (Reviewed in  
10 Aggarwal, M. L. et al. (1998) J. Biol. Chem. 273:1-4; Levine, A. (1997) Cell 88:323-331.) About 50% of all human cancers contain mutations in the *p53* gene. These mutations result in either the absence of functional p53 or, more commonly, a defective form of p53 which is overexpressed. p53 is a transcription factor that contains a central core domain required for DNA binding. Most cancer-associated mutations in p53 localize to this domain. In normal proliferating cells, p53 is expressed at  
15 low levels and is rapidly degraded. p53 expression and activity is induced in response to DNA damage, abortive mitosis, and other stressful stimuli. In these instances, p53 induces apoptosis or arrests cell growth until the stress is removed. Downstream effectors of p53 activity include apoptosis-specific proteins and cell cycle regulatory proteins, including Rb, oncogene products, cyclins, and cell cycle-dependent kinases.

20 The metastasis-suppressor gene KAI1 (CD82) has been reported to be related to the tumor suppressor gene p53. KAI1 is involved in the progression of human prostatic cancer and possibly lung and breast cancers when expression is decreased. KAI1 encodes a member of a structurally distinct family of leukocyte surface glycoproteins. The family is known as either the tetraspan transmembrane protein family or transmembrane 4 superfamily (TM4SF) as the members of this  
25 family span the plasma membrane four times. The family is composed of integral membrane proteins having a N-terminal membrane-anchoring domain which functions as both a membrane anchor and a translocation signal during protein biosynthesis. The N-terminal membrane-anchoring domain is not cleaved during biosynthesis. TM4SF proteins have three additional transmembrane regions, seven or more conserved cysteine residues, are similar in size (218 to 284 residues), and all have a large  
30 extracellular hydrophilic domain with three potential N-glycosylation sites. The promoter region contains many putative binding motifs for various transcription factors, including five AP2 sites and nine Sp1 sites. Gene structure comparisons of KAI1 and seven other members of the TM4SF indicate that the splicing sites relative to the different structural domains of the predicted proteins are conserved. This suggests that these genes are related evolutionarily and arose through gene  
35 duplication and divergent evolution (Levy, S. et al. (1991) J. Biol. Chem. 266:14597-14602; Dong,

J.T. et al. (1995) *Science* 268:884-886; Dong, J.T. et al., (1997) *Genomics* 41:25-32).

The Leucine-rich gene-Glioma Inactivated (LGI1) protein shares homology with a number of transmembrane and extracellular proteins which function as receptors and adhesion proteins. LGI1 is encoded by an LLR (leucine-rich, repeat-containing) gene and maps to 10q24. LGI1 has four LLRs  
5 which are flanked by cysteine-rich regions and one transmembrane domain (Somerville, R.P. et al. (2000) *Mamm. Genome* 11:622-627). LGI1 expression is seen predominantly in neural tissues, especially brain. The loss of tumor suppressor activity is seen in the inactivation of the LGI1 protein which occurs during the transition from low to high-grade tumors in malignant gliomas. The reduction of LGI1 expression in low grade brain tumors and its significant reduction or absence of  
10 expression in malignant gliomas suggests that it could be used for diagnosis of glial tumor progression (Chernova, O.B. et al. (1998) *Oncogene* 17:2873-2881).

The ST13 tumor suppressor was identified in a screen for factors related to colorectal carcinomas by subtractive hybridization between cDNA of normal mucosal tissues and mRNA of colorectal carcinoma tissues (Cao, J. et al. (1997) *J. Cancer Res. Clin. Oncol.* 123:447-451). ST13 is  
15 down-regulated in human colorectal carcinomas.

Mutations in the von Hippel-Lindau (VHL) tumor suppressor gene are associated with retinal and central nervous system hemangioblastomas, clear cell renal carcinomas, and pheochromocytomas (Hoffman, M. et al. (2001) *Hum. Mol. Genet.* 10:1019-1027; Kamada, M. (2001) *Cancer Res.* 61:4184-4189). Tumor progression is linked to defects or inactivation of the VHL gene. VHL  
20 regulates the expression of transforming growth factor- $\alpha$ , the GLUT-1 glucose transporter and vascular endothelial growth factor. The VHL protein associates with elongin B, elongin C, Cul2 and Rbx1 to form a complex that regulates the transcriptional activator hypoxia-inducible factor (HIF). HIF induces genes involved in angiogenesis such as vascular endothelial growth factor and platelet-derived growth factor B. Loss of control of HIF caused by defects in VHL results in the excessive  
25 production of angiogenic peptides. VHL may play roles in inhibition of angiogenesis, cell cycle control, fibronectin matrix assembly, cell adhesion, and proteolysis.

Mutations in tumor suppressor genes are a common feature of many cancers and often appear to affect a critical step in the pathogenesis and progression of tumors. Accordingly, Chang, F. et al. (1995; *J. Clin. Oncol.* 13:1009-1022) suggest that it may be possible to use either the gene or an  
30 antibody to the expressed protein 1) to screen patients at increased risk for cancer, 2) to aid in diagnosis made by traditional methods, and 3) to assess the prognosis of individual cancer patients. In addition, Hamada, K. et al. (1996; *Cancer Res.* 56:3047-3054) are investigating the introduction of p53 into cervical cancer cells via an adenoviral vector as an experimental therapy for cervical cancer.

The PR-domain genes were recently recognized as playing a role in human tumorigenesis.  
35 PR-domain genes normally produce two protein products: the PR-plus product, which contains the PR

domain, and the PR-minus product which lacks this domain. In cancer cells, PR-plus is disrupted or overexpressed, while PR-minus is present or overexpressed. The imbalance in the amount of these two proteins appears to be an important cause of malignancy (Jiang, G.L. and S. Huang (2000) *Histol. Histopathol.* 15:109-117).

5 Many neoplastic disorders in humans can be attributed to inappropriate gene transcription. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes (Cleary, M.L. (1992) *Cancer Surv.* 15:89-104). Chromosomal translocations may also produce chimeric loci which fuse the coding sequence of one gene with the regulatory regions of a second unrelated gene. An important class of transcriptional  
10 regulators are the zinc finger proteins. The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern include the C2H2-type, C4-type, and C3HC4-type zinc fingers, and the PHD domain (Lewin, B. (1990) *Genes IV*, Oxford University Press, New York, NY, and Cell Press, Cambridge, MA, pp. 554-570; Aasland, R., et al. (1995) *Trends Biochem.*  
15 *Sci.* 20:56-59). One clinically relevant zinc-finger protein is WT1, a tumor-suppressor protein that is inactivated in children with Wilm's tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, A.G. (1995) *N. Engl. J. Med.* 332:45-47).

#### Tumor responsive proteins

20 Cancers, also called neoplasias, can be divided into three categories: carcinomas, sarcomas, and leukemias. Carcinomas are malignant growths of soft epithelial cells that may infiltrate surrounding tissues and give rise to metastatic tumors. Sarcomas may be of epithelial origin or arise from connective tissue. Leukemias are progressive malignancies of blood-forming tissue characterized by proliferation of leukocytes and their precursors, and may be classified as  
25 myelogenous (granulocyte- or monocyte-derived) or lymphocytic (lymphocyte-derived). Tumorigenesis refers to the progression of a tumor's growth from its inception. Malignant cells may be quite similar to normal cells within the tissue of origin or may be undifferentiated (anaplastic). Tumor cells may possess few nuclei or one large polymorphic nucleus. Anaplastic cells may grow in a disorganized mass that is poorly vascularized and as a result contains large areas of ischemic  
30 necrosis. Differentiated neoplastic cells may secrete the same proteins as the tissue of origin. Cancers grow, infiltrate, invade, and destroy the surrounding tissue through direct seeding of body cavities or surfaces, through lymphatic spread, or through hematogenous spread. Cancer remains a major public health concern and current preventative measures and treatments do not match the needs of most patients. Understanding of the neoplastic process of tumorigenesis can be aided by the  
35 identification of molecular markers of prognostic and diagnostic importance.

Current forms of cancer treatment include the use of immunosuppressive drugs (Morisaki, T. et al. (2000) *Anticancer Res.* 20:3363-3373; Geoerger, B. et al. (2001) *Cancer Res.* 61:1527-1532). The identification of proteins involved in cell signaling, and specifically proteins that act as receptors for immunosuppressant drugs, may facilitate the development of anti-tumor agents. For example, immunophilins are a family of conserved proteins found in both prokaryotes and eukaryotes that bind to immunosuppressive drugs with varying degrees of specificity. One such group of immunophilic proteins is the peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) family (PPIase, rotamase). These enzymes, first isolated from porcine kidney cortex, accelerate protein folding by catalyzing the cis-trans isomerization of proline imidic peptide bonds in oligopeptides (Fischer, G. and F.X. Schmid (1990) *Biochemistry* 29:2205-2212). Included within the immunophilin family are the cyclophilins (e.g., peptidyl-prolyl isomerase A or PPIA) and FK-binding protein (e.g., FKBP) subfamilies. Cyclophilins are multifunctional receptor proteins which participate in signal transduction activities, including those mediated by cyclosporin (or cyclosporine). The PPIase domain of each family is highly conserved between species. Although structurally distinct, these multifunctional receptor proteins are involved in numerous signal transduction pathways, and have been implicated in folding and trafficking events.

The immunophilin protein cyclophilin binds to the immunosuppressant drug cyclosporin A. FKBP, another immunophilin, binds to FK506 (or rapamycin). Rapamycin is an immunosuppressant agent that arrests cells in the G<sub>1</sub> phase of growth, inducing apoptosis. Like cyclophilin, this macrolide antibiotic (produced by *Streptomyces tsukubaensis*) acts by binding to ubiquitous, predominantly cytosolic immunophilin receptors. These immunophilin/immunosuppressant complexes (e.g., cyclophilin A/cyclosporin A (CypA/CsA) and FKBP12/FK506) achieve their therapeutic results through inhibition of the phosphatase calcineurin, a calcium/calmodulin-dependent protein kinase that participates in T-cell activation (Hamilton, G.S. and J.P. Steiner (1998) *J. Med. Chem.* 41: 5119-5143). The murine *fkbp51* gene is abundantly expressed in immunological tissues, including the thymus and T lymphocytes (Baughman, G. et al. (1995) *Molec. Cell. Biol.* 15: 4395-4402). FKBP12/rapamycin-directed immunosuppression occurs through binding to TOR (yeast) or FRAP (FKBP12-rapamycin-associated protein, in mammalian cells), the kinase target of rapamycin essential for maintaining normal cellular growth patterns. Dysfunctional TOR signaling has been linked to various human disorders including cancer (Metcalf, S.M. et al. (1997) *Oncogene* 15:1635-1642; Emami, S. et al. (2001) *FASEB J.* 15:351-361), and autoimmunity (Damoiseaux, J.G. et al. (1996) *Transplantation* 62:994-1001).

Several cyclophilin isozymes have been identified, including cyclophilin B, cyclophilin C, mitochondrial matrix cyclophilin, bacterial cytosolic and periplasmic PPIases, and natural-killer cell cyclophilin-related protein possessing a cyclophilin-type PPIase domain, a putative tumor-recognition



complex involved in the function of natural killer (NK) cells. These cells participate in the innate cellular immune response by lysing virally-infected cells or transformed cells. NK cells specifically target cells that have lost their expression of major histocompatibility complex (MHC) class I genes (common during tumorigenesis), endowing them with the potential for attenuating tumor growth. A  
5 150-kDa molecule has been identified on the surface of human NK cells that possesses a domain which is highly homologous to cyclophilin/peptidyl-prolyl cis-trans isomerase. This cyclophilin-type protein may be a component of a putative tumor-recognition complex, a NK tumor recognition sequence (NK-TR) (Anderson, S.K. et al. (1993) Proc. Natl. Acad. Sci. USA 90:542-546). The NKTR tumor recognition sequence mediates recognition between tumor cells and large granular  
10 lymphocytes (LGLs), a subpopulation of white blood cells (comprised of activated cytotoxic T cells and natural killer cells) capable of destroying tumor targets. The protein product of the NKTR gene presents on the surface of LGLs and facilitates binding to tumor targets. More recently, a mouse Nktr gene and promoter region have been located on chromosome 9. The gene encodes a NK-cell-specific 150-kDa protein (NK-TR) that is homologous to cyclophilin and other tumor-responsive proteins  
15 (Simons-Evelyn, M. et al. (1997) Genomics 40:94-100).

Other proteins that interact with tumorigenic tissue include cytokines such as tumor necrosis factor (TNF). The TNF family of cytokines are produced by lymphocytes and macrophages, and can cause the lysis of transformed (tumor) endothelial cells. Endothelial protein 1 (Edp1) has been identified as a human gene activated transcriptionally by TNF-alpha in endothelial cells, and a TNF-alpha inducible Edp1 gene has been identified in the mouse (Swift, S. et al. (1998) Biochim. Biophys. Acta 1442:394-398).

#### Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support.  
25 Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the  
30 expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic  
35 predisposition, condition, disease, or disorder.

Adipocyte maturation

The primary function of adipose tissue is the ability to store and release fat during periods of feeding and fasting. White adipose tissue is the major energy reserve in periods of fasting, and its reserve is mobilized during energy deprivation. Adipose tissue is one of the primary target tissues for insulin, and adipogenesis and insulin resistance are linked in type II diabetes, non-insulin dependent diabetes mellitus (NIDDM). Cytologically the conversion of a preadipocytes into mature adipocytes is characterized by deposition of fat droplets around the nuclei. The conversion process *in vivo* can be induced by thiazolidinediones and other PPAR $\gamma$  agonists (Adams et al. (1997) J Clin Invest 100:3149-3153) which also lead to increased sensitivity to insulin and reduced plasma glucose and blood pressure.

Pickup and Crook (1998; Diabetologia 41:1241-8) have suggested that NIDDM may result from the inability of an individual with hypersensitive acute-phase immune response to carry out normal cell signaling and repair. Steps in this process are highly correlated with long-term lifestyle and environment and include: 1) high glucose stimulation of insulin and cytokine production, 2) influence of various cytokines on tissue remodeling during adipocyte differentiation and their affect on signaling pathways, and 3) occurrence of tissue damage when cytokines continue to be produced, extracellular matrix components (ECM) are not recycled, and homeostasis is not timely restored. Many cytokines and the receptors with which they interact are implicated in this process. These cytokines include tumor necrosis factor, connective tissue growth factor, transforming growth factor-beta, interleukin (IL)-13 and their receptors. Tumor necrosis factor contributes to insulin resistance by inhibiting insulin-stimulated tyrosine phosphorylation of the insulin receptor. This, in turn, prevents the insulin receptor from participating in normal signaling processes (Skolnik and Marcusohn (1996) Cytokine Growth Factor Rev 7:161-173; Hotamisligil (1999) J Intern med 245:621-625). Connective tissue growth factor mediates the buildup of mesengial matrix (Murphy et al. (2000) J Biol Chem 274:5830-5834). Transforming growth factor-beta mediates the buildup of mesengial matrix of the kidney and affects vascular function through its interaction with the inositol trisphosphate receptor, a key intracellular calcium channel (Sharma and McGowan (2000) Cytokine Growth Factor Rev 11:115-123).

IL-13 and IL-4 are immuno-regulatory cytokines which share many overlapping biological properties. They both promote growth of B-cells (McKenzie et al. (1993) Proc Natl Acad Sci 90:3735-3739), induce expression of germ line C $\epsilon$  transcripts, and direct naive B cells to switch to the synthesis of IgE and IgG4 (Punnomen et al. (1993) Proc Natl Acad Sci 90:3730-3734). Similarly, different isoforms of the IL-13 and IL-4 receptors interact to form four types of IL-13 receptor complexes. In some instances, IL-13 utilizes a receptor complex composed of the IL-4 receptor- $\alpha$  chain (R $\alpha$ ) and the IL-13R $\alpha$ . Although the specific role of each chain in IL-13 signaling is unclear,

Ba/F3 cells transfected with IL-13R $\alpha$ 1 display a mitogenic response to IL-13, but cells transfected with mouse IL-13R $\alpha$ 2 do not. In addition, a soluble IL-13R $\alpha$ 2/Fc fusion protein blocks the mitogenic response to IL-13 (Donaldson et al. (1998) J Immunol 161:2317-2324). This suggests that IL-13R $\alpha$ 2 could serve as a dominant negative inhibitor or decoy receptor for IL-13. However, in colonic carcinoma cell lines, the receptor complex displayed growth inhibition which was associated with tyrosine phosphorylation of insulin receptor substrate-1. It is evident that more research is needed to establish 1) which isoforms of the receptor complex promote cell growth and which inhibit cell growth and 2) whether this varies by cell or tissue type.

Most adipocyte research has been carried out using mouse cell lines. Recent evidence, however, indicates that culture conditions which stimulate mouse preadipocyte differentiation are different from those which induce human preadipocytes. In addition to the known genetic differences between these species, diploid human primary cells respond differently than aneuploid mouse cells.

#### Breast Cancer

There are more than 180,000 new cases of breast cancer diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (Gish, K. (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).

Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, *supra*). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

The relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie, K. et al. (1993) Cancer and Metastasis Rev. 12:255-274, and references cited therein for a review of this area.)

Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR

is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S. S. et al. (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a  
5 human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down regulated in mammary carcinoma cells relative to normal mammary epithelial cells  
10 (Zhou, Z. et al. (1998) Int. J. Cancer 78:95-99; Chen, L. et al. (1990) Oncogene 5:1391-1395; Ulrix, W. et al (1999) FEBS Lett 455:23-26; Sager, R. et al. (1996) Curr. Top. Microbiol. Immunol. 213:51-64; and Lee, S. W. et al. (1992) Proc. Natl. Acad. Sci. USA 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it  
15 has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba, I.I. et al. (1998) Clin. Cancer Res. 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

#### Prostate Cancer

20 Prostate cancer is a common malignancy in men over the age of 50, and the incidence increases with age. In the US, there are approximately 132,000 newly diagnosed cases of prostate cancer and more than 33,000 deaths from the disorder each year.

Once cancer cells arise in the prostate, they are stimulated by testosterone to a more rapid growth. Thus, removal of the testes can indirectly reduce both rapid growth and metastasis of the  
25 cancer. Over 95 percent of prostatic cancers are adenocarcinomas which originate in the prostatic acini. The remaining 5 percent are divided between squamous cell and transitional cell carcinomas, both of which arise in the prostatic ducts or other parts of the prostate gland.

As with most cancers, prostate cancer develops through a multistage progression ultimately resulting in an aggressive, metastatic phenotype. The initial step in tumor progression involves the  
30 hyperproliferation of normal luminal and/or basal epithelial cells that become hyperplastic and evolve into early-stage tumors. The early-stage tumors are localized in the prostate but eventually may metastasize, particularly to the bone, brain or lung. About 80% of these tumors remain responsive to androgen treatment, an important hormone controlling the growth of prostate epithelial cells. However, in its most advanced state, cancer growth becomes androgen-independent and there is  
35 currently no known treatment for this condition.

A primary diagnostic marker for prostate cancer is prostate specific antigen (PSA). PSA is a tissue-specific serine protease almost exclusively produced by prostatic epithelial cells. The quantity of PSA correlates with the number and volume of the prostatic epithelial cells, and consequently, the levels of PSA are an excellent indicator of abnormal prostate growth. Men with prostate cancer exhibit an early linear increase in PSA levels followed by an exponential increase prior to diagnosis. However, since PSA levels are also influenced by factors such as inflammation, androgen and other growth factors, some scientists maintain that changes in PSA levels are not useful in detecting individual cases of prostate cancer.

Current areas of cancer research provide additional prospects for markers as well as potential therapeutic targets for prostate cancer. Several growth factors have been shown to play a critical role in tumor development, growth, and progression. The growth factors Epidermal Growth Factor (EGF), Fibroblast Growth Factor (FGF), and Tumor Growth Factor alpha (TGF $\alpha$ ) are important in the growth of normal as well as hyperproliferative prostate epithelial cells, particularly at early stages of tumor development and progression, and affect signaling pathways in these cells in various ways (Lin J *et al.* (1999) Cancer Res. 59:2891-2897; Putz T *et al.* (1999) Cancer Res 59:227-233). The TGF- $\beta$  family of growth factors are generally expressed at increased levels in human cancers and the high expression levels in many cases correlates with advanced stages of malignancy and poor survival (Gold LI (1999) Crit Rev Oncog 10:303-360). Finally, there are human cell lines representing both the androgen-dependent stage of prostate cancer (LNCap) as well as the androgen-independent, hormone refractory stage of the disease (PC3 and DU-145) that have proved useful in studying gene expression patterns associated with the progression of prostate cancer, and the effects of cell treatments on these expressed genes (Chung TD (1999) Prostate 15:199-207).

#### Lung Cancer

Lung cancer is the leading cause of cancer death for men and the second leading cause of cancer death for women in the U.S. The vast majority of lung cancer cases are attributed to smoking tobacco, and increased use of tobacco products in third world countries is projected to lead to an epidemic of lung cancer in these countries. Exposure of the bronchial epithelium to tobacco smoke appears to result in changes in tissue morphology, which are thought to be precursors of cancer. Lung cancers are divided into four histopathologically distinct groups. Three groups (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) are classified as non-small cell lung cancers (NSCLCs). The fourth group of cancers is referred to as small cell lung cancer (SCLC). Collectively, NSCLCs account for ~70% of cases while SCLCs account for ~18% of cases. The molecular and cellular biology underlying the development and progression of lung cancer are incompletely understood.

Deletions on chromosome 3 are common in this disease and are thought to indicate the

presence of a tumor suppressor gene in this region. Activating mutations in K-ras are commonly found in lung cancer and are the basis of one of the mouse models for the disease.

#### Leukemias

Leukemias can be classified into four major categories, and all involve malignant transformation of pluripotent stem cells. Acute leukemias, both lymphoblastic (ALL) and myeloid (AML) types, are characterized by the presence of immature cells in the blood. Chronic leukemias, both lymphocytic (CLL) and myelocytic (CML), are associated with mature, differentiated cells, but proportions of each cell type are abnormal. For example, CLL patients usually have clonal expansion of B cell lymphocytes. CML patients often have granulocytes of all stages of maturity present in blood, bone marrow, and other organs. Monoclonal antibodies specific for B- and T-cells are helpful diagnostic tools, in addition to histological analysis. Disease progresses as normal hematopoietic bone marrow is displaced by malignant cells. Cause has been determined to be genetic in some cases, and chemical or radiation-induced in others.

#### Genes Regulated in Dendritic Cell Differentiation

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for examining which genes are tissue specific, carrying out housekeeping functions, parts of a signaling cascade, or specifically related to a particular genetic predisposition, condition, disease, or disorder.

The potential application of gene expression profiling is particularly relevant to characterizing lineage differences during cellular development that will improve diagnosis, prognosis, and treatment of disease. For example, both the levels and sequences expressed in dendritic cells from subjects with autoimmunity may be compared with the levels and sequences expressed in dendritic cells from normal subjects.

Dendritic cells (DC) are antigen presenting cells (APC) that play a key role in the primary immune response because of their unique ability to present antigens to naive T cells. In addition, DC differentiate into separate subsets that sustain and regulate immune responses following initial contact with antigen. DC subsets include those that preferentially induce particular T helper 1 (Th1) or T helper 2 (Th2) responses and those that regulate B cell responses. Moreover, DC are increasingly being used to manipulate immune responses, either to downregulate an aberrant autoimmune response or to enhance vaccination or a tumor-specific response.

DC are functionally specialized in correlation with their particular differentiation state. CD34+ myeloid cells found in the bone marrow mature in response to as yet unclear signals into CD14+ CD11c+ monocytes. An innate or antigen non-specific response takes place initially when

monocytes circulate to nonlymphoid tissues and respond to lipopolysaccharide (LPS), a bacterially-derived mitogen, and viruses. Such direct encounter with antigen causes secretion of pro-inflammatory cytokines that attract and regulate natural killer cells, macrophages, and eosinophils in the first line of defense against invading pathogens. Monocytes then mature into DC, which capture  
5 antigen highly efficiently through endocytosis and antigen receptor uptake. Antigen processing and presentation trigger activation and differentiation into mature DC that express MHC class II molecules on the cell surface and efficiently activate T cells, initiating antigen-specific T cell and B cell responses. In turn, T cells activate DC through CD40 ligand - CD40 interactions, which stimulate expression of the costimulatory molecules CD80 and CD86, the latter most potent in amplifying T  
10 cell responses. DC interaction via CD40 with T cells also stimulates the production of inflammatory cytokines such as TNF alpha and IL-1. Engagement of RANK, a member of the TNF receptor family by its ligand, TRANCE, which is expressed on activated T cells, enhances the survival of DC through inhibition of apoptosis, thereby enhancing T cell activation. The maturation and differentiation of monocytes into mature DC links the antigen non-specific innate immune response to the antigen-  
15 specific adaptive immune response.

The process by which monocytes differentiate into immature dendritic cells in vivo has not been fully elucidated. Incubation of monocytes with granulocyte-macrophage colony stimulating factor (GM-CSF) and interleukin (IL) -4 in vitro yields cells that exhibit functional and morphological characteristics equivalent to immature dendritic cells found in vivo. Moreover, incubation in vitro of  
20 immature dendritic cells with tumor necrosis factor alpha (TNF- $\alpha$ ), CD40 ligand, LPS, or monocyte-conditioned medium yields mature dendritic cells that are potent activators of naive T cells.

The ability to manipulate DC in vitro and their capacity to mount an effective immune response with small numbers of DC and little antigen has led to potential immunotherapies for diseases such as cancer, AIDS, and infectious diseases; and enhancing vaccine efficacy. Spontaneous  
25 remissions of particular cancers such as renal cell carcinomas and melanomas indicate that the immune system can respond to tumor antigens and eliminate tumors. However, tumors escape immune surveillance through a number of means including secretion of IL-10, macrophage colony stimulating factor, IL-6, and vascular endothelial growth factor, all of which inhibit DC activity and promote tolerance of tumor tissue. Delivery of tumor antigen-loaded DC to tumors can induce tumor-  
30 specific rejection in animal models. Similarly, pathogens can escape immune surveillance by altering antigen processing and presentation pathways or interfering with maturation of antigen presenting cells. Rather than providing resistance, DC can complicate infection by hosting latent viruses such as Kaposi's virus and cytomegalovirus, complicating infection. HIV-1 and measles virus particles are efficiently produced in DC. Vaccines against tumors or infectious pathogens could be improved by  
35 systemic or local administration of DC loaded with tumor antigens or attenuated viral particles or

components, respectively.

The expression of killer-inhibitor regulatory molecules, chemokines, chemokine receptors, and proteinases have been identified in DC through sequencing of ESTs. Continuing this search may reveal new lymphocyte-binding and antigen-processing molecules, transmembrane and secretory products, and transcription factors that may help to explain the specialized features of DC and allow manipulation of the immune system.

#### Tumor progression

Endometrial cancer is the most common gynecologic cancer. Approximately 90% of endometrial cancers are epithelial in origin, and 90% of these cancers are classified as endometrial adenocarcinomas. Estrogen appears to act as a tumor promoter in endometrial tissue. Evidence indicates that p53 and Ki-ras are mutated in endometrial cancer. However, these mutations occur in a small percentage of cases and do not appear to be the initiating events in the disease. In addition, most chromosomes contain regions of allelic loss in endometrial cancer, indicating that many genes may be affected in this disease.

#### Biopharmacological tools for microarray analysis

Human umbilical vein endothelial cells (HUVECs) are a primary cell line derived from the endothelium of the human umbilical vein. HUVECs are used to study the functional biology of human endothelial cells *in vitro*. Activation of vascular endothelium is observed in physiological and pathophysiological processes including vascular tone regulation, coagulation and thrombosis, atherosclerosis, and inflammation.

Tumor necrosis factor-alpha (TNF- $\alpha$ ) is a pleiotropic cytokine that plays a central role in mediation of the inflammatory response through activation of multiple signal transduction pathways. TNF- $\alpha$  is produced by activated lymphocytes, macrophages, and other white blood cells, and activates endothelial cells.

PMA is an agonist of protein kinase C (PKC). PKC is a calcium-activated, phospholipid-dependent serine- and threonine-specific kinase that, upon activation, phosphorylates a broad range of secondary targets. TNF- $\alpha$  causes translocation of PKC from the cytosol to the membrane where it phosphorylates a variety of targets.

Interleukin 1 beta (IL-1 $\beta$ ) is a cytokine associated with acute inflammatory responses and is involved in processes such as fever induction, metabolic regulation, and bone remodeling. Both cells of the immune system (monocytes, dendritic cells, NK cells, platelets, and neutrophils) and somatic cells (osteoblasts, neurons, Schwann cells, oligodendrocytes, and adrenal cortical cells) can produce IL-1 $\beta$ . IL-1 $\beta$  can induce its own production in monocytes, the production of adhesion molecules and chemokines in endothelial cells, and interferon gamma (IFN- $\gamma$ ) production by NK cells in conjunction with IL-12. IL-1 is produced as a single chain pro-molecule that must be cleaved by a specialized



protease – IL-1 Converting Enzyme (ICE) – to acquire its function. Interleukin 10 (IL-10) is produced by CD4+ T cell clones and some CD8+ T cell clones. Human B cells, EBV-transformed lymphoblastoid cell lines, and monocytes can also produce IL-10 upon activation. IL-10 is a pleiotrophic cytokine that can exert either immunostimulatory or immunosuppressive effects on a variety of cell types. It is a potent immunosuppressant of macrophage functions. In vitro, IL-10 can inhibit the accessory function and antigen-presenting capacity of monocytes by, among other effects, downregulating class II MHC expression. Thus, IL-10 can inhibit monocyte/macrophage-dependent, antigen specific proliferation of mouse Th1 clones as well as human Th0-, Th1-, and Th2- like T cells. IL-10 can also inhibit the monocyte/macrophage-dependent, antigen stimulated cytokine synthesis (especially IFN-g) by human PBMNC and NK. Additionally, IL-10 is a potent inhibitor of monocyte/macrophage activation and its resultant cytotoxic effects. It can suppress the production of numerous cytokines including TNF-a, IL-1, IL-6, and IL-10, as well as the synthesis of superoxide anion, reactive oxygen intermediates, and reactive nitrogen intermediates by activated monocytes/macrophages. As an immunostimulatory cytokine, IL-10 can act on B cells to enhance their viability, cell proliferation, Ig secretion, and class II MHC expression. Aside from B-lymphocytes, IL-10 is also a growth co-stimulator for thymocytes and mast cells, as well as an enhancer of cytotoxic T cell development.

Thiazolidinediones or peroxisome proliferator-activated receptor  $\gamma$  (PPAR- $\gamma$ ) agonists are a new class of antidiabetic agents that improve insulin sensitivity and reduce plasma glucose and blood pressure in subjects with type II diabetes. These agents can bind and activate an orphan nuclear receptor and some of them induce human adipocyte differentiation.

#### RNA Expression

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for examining which genes are tissue specific, carrying out housekeeping functions, parts of a signaling cascade, or specifically related to a particular genetic predisposition, condition, disease, or disorder.

The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of disease. For example, both the levels and sequences expressed in tissues from subjects with a cardiovascular disorder may be compared with the levels and sequences expressed in normal tissue.

Atherosclerosis and the associated coronary artery disease and cerebral stroke represent the most common cause of death in industrialized nations. Although certain key risk factors have been identified, a full molecular characterization that elucidates the causes and provide care for this

complex disease has not been achieved. Molecular characterization of growth and regression of atherosclerotic vascular lesions requires identification of the genes that contribute to features of the lesion including growth, stability, dissolution, rupture and, most lethally, induction of occlusive vessel thrombus. Vascular lesions principally involve the vascular endothelium and the surrounding  
5 smooth muscle tissue.

Development of atherosclerosis is understood to be induced by the presence of circulating lipoprotein. Lipoproteins, such as the cholesterol-rich low-density lipoprotein (LDL), accumulate in the extracellular space of the vascular intima, and undergo modification. Oxidation of LDL (Ox-LDL) occurs most avidly in the sub-endothelial space where circulating antioxidant defenses are less  
10 effective. Mononuclear phagocytes enter the intima, differentiate into macrophages, and ingest modified lipids including Ox-LDL. During Ox-LDL uptake, macrophages produce cytokines (e.g. tumor necrosis factor  $\alpha$  (TNF- $\alpha$ ) and interleukin-1 (IL-1)) and growth factors (e.g. M-CSF, VEGF, and PDGF-BB) that elicit further cellular events that modulate atherogenesis such as smooth muscle cell proliferation and production of extracellular matrix by vascular endothelium. Additionally, these  
15 macrophages may activate genes in endothelium and smooth muscle tissue involved in inflammation and tissue differentiation, including superoxide dismutase (SOD), IL-8, and ICAM-1.

The vascular endothelium influences not only the three classically interacting components of hemostasis: the vessel, the blood platelets and the clotting and fibrinolytic systems of plasma, but also the natural sequelae: inflammation and tissue repair. Two principal modes of endothelial  
20 behavior may be differentiated, best defined as an anti- and a prothrombotic state. Under physiological conditions endothelium mediates vascular dilatation (formation of nitric oxide (NO), PGI<sub>2</sub>, adenosine, hyperpolarising factor), prevents platelet adhesion and activation (production of adenosine, NO and PGI<sub>2</sub>, removal of ADP), blocks thrombin formation (tissue factor pathway inhibitor, activation of protein C via thrombomodulin, activation of antithrombin III) and mitigates  
25 fibrin deposition (t- and scu plasminogen activator production). Adhesion and transmigration of inflammatory leukocytes are attenuated, e.g. by NO and IL-10, and oxygen radicals are efficiently scavenged (urate, NO, glutathione, SOD).

When the endothelium is physically disrupted or functionally perturbed by postischemic reperfusion, acute and chronic inflammation, atherosclerosis, diabetes and chronic arterial  
30 hypertension, then completely opposing actions pertain. This prothrombotic, proinflammatory state is characterised by vaso-constriction, platelet and leukocyte activation and adhesion (externalisation, expression and upregulation of, for example, von Willebrand factor, platelet activating factor, P-selectin, ICAM-1, IL-8, MCP-1, and TNF- $\alpha$ ), promotion of thrombin formation, coagulation and fibrin deposition at the vascular wall (expression of tissue factor, PAI-1, and phosphatidyl serine) and,  
35 in platelet-leukocyte coaggregates, additional inflammatory interactions via attachment of platelet

CD40-ligand to endothelial, monocyte and B-cell CD40. Since thrombin formation and inflammatory stimulation set the stage for later tissue repair, complete abolition of such endothelial responses cannot be the goal of clinical interventions aimed at limiting procoagulatory, prothrombotic actions of a dysfunctional vascular endothelium. (See, e.g., Becker et al. (2000) *Z Kardiol* 89:160-167.)

5 Tumor necrosis factor  $\alpha$  is a pleiotropic cytokine that mediates immune regulation and inflammatory responses. TNF- $\alpha$ -related cytokines generate partially overlapping cellular responses, including differentiation, proliferation, nuclear factor- $\kappa$ B (NF- $\kappa$ B) activation, and cell death, by triggering the aggregation of receptor monomers (Smith, C.A. et al. (1994) *Cell* 76:959-962). The cellular responses triggered by TNF- $\alpha$  are initiated through its interaction with distinct cell surface  
10 receptors (TNFRs). NF- $\kappa$ B is a transcription factor with a pivotal role in inducing genes involved in physiological processes as well as in the response to injury and infection. Activation of NF- $\kappa$ B involves the phosphorylation and subsequent degradation of an inhibitory protein, I $\kappa$ B, and many of the proximal kinases and adaptor molecules involved in this process have been elucidated. Additionally, the NF- $\kappa$ B activation pathway from cell membrane to nucleus for IL-1 and TNF- $\alpha$  is  
15 now understood (Bowie and O'Neill (2000) *Biochem Pharmacol* 59:13-23).

Monocyte chemoattractant protein-1 (MCP-1) is known to play an important role in the pathogenesis of atherosclerosis by inducing monocyte migration. TNF- $\alpha$  treatment of human umbilical vein endothelial cells (HUVECs) increased the cellular secretions of MCP-1 119-fold compared with untreated cells. Troglitazone, an insulin-sensitizing drug, significantly inhibited this  
20 TNF- $\alpha$ -induced increase in MCP-1 secretions and decreased mRNA levels (Ohta et al. (2000) *Diabetes Res Clin Pract* 48:171-176).

Treatment of confluent cultures of vascular smooth muscle cells (SMCs) with TNF- $\alpha$  suppresses the incorporation of [ $^3$ H]proline into both collagenase-digestible proteins (CDP) and noncollagenous proteins (NCP). Such suppression by TNF- $\alpha$  is not observed in confluent bovine  
25 aortic endothelial cells and human fibroblastic IMR-90 cells. TNF- $\alpha$  decreases the relative proportion of collagen types IV and V suggesting that TNF- $\alpha$  modulates collagen synthesis by SMCs depending on their cell density and therefore may modify formation of atherosclerotic lesions (Hiraga et al. (2000) *Life Sci* 66:235-244).

Human coronary artery smooth muscle cells (CASMC) are primary cells isolated from the  
30 tunica media (an intermediate muscular layer) of a human coronary artery. Vascular smooth muscle cells are a model of increasing significance in vascular biology. It is now well known that besides their obvious role in the regulation of vascular tone and, consequently, oxygen supply to various tissues, their behavior under inflammatory conditions is an important factor in the development of atherosclerosis and restenosis.

35 Human aortic endothelial cells (HAECs) are primary cells derived from the endothelium of a

human aorta. HAECs have been used as an experimental model for investigating *in vitro* the role of the endothelium in human vascular biology. Activation of the vascular endothelium is considered to be a central event in a wide range of both physiological and pathophysiological processes, such as vascular tone regulation, coagulation and thrombosis, atherosclerosis, and inflammation.

5           Thus, vascular tissue genes differentially expressed during treatment of CASMC and HAEC cell cultures with TNF $\alpha$  may reasonably be expected to be markers of the atherosclerotic process.

#### Steroid Hormones

Glucocorticoids are naturally occurring hormones that prevent or suppress inflammation and immune responses when administered at pharmacological doses. At the molecular level, unbound  
10 glucocorticoids readily cross cell membranes and bind with high affinity to specific cytoplasmic receptors. Subsequent to binding, transcription and, ultimately, protein synthesis are affected. The result can include inhibition of leukocyte infiltration at the site of inflammation, interference in the function of mediators of inflammatory response, and suppression of humoral immune responses. The antiinflammatory actions of corticosteroids are thought to involve phospholipase A2 inhibitory  
15 proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of the precursor molecule arachidonic acid. Further, corticosteroids inhibit eosinophil, basophil, and airway epithelial cell function by regulation of cytokines that mediate the inflammatory response. They inhibit leukocyte infiltration at the site of inflammation, interfere in the function of mediators of the  
20 inflammatory response, and suppress the humoral immune response. Corticosteroids are used to treat allergies, asthma, arthritis, and skin conditions. Beclomethasone is a synthetic glucocorticoid that is used to treat steroid-dependent asthma, to relieve symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or to prevent recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive effects of intranasal beclomethasone are 5000 times greater than  
25 those produced by hydrocortisone.

#### Colon Cancer

Colorectal cancer is the second leading cause of cancer deaths in the United States. Colon cancer is associated with aging, since 90% of the total cases occur in individuals over the age of 55. A widely accepted hypothesis is that several contributing genetic mutations must accumulate over  
30 time in an individual who develops the disease. To understand the nature of genetic alterations in colorectal cancer, a number of studies have focused on the inherited syndromes. The first known inherited syndrome, Familial Adenomatous Polyposis (FAP), is caused by mutations in the Adenomatous Polyposis Coli gene (APC), resulting in truncated or inactive forms of the protein. This tumor suppressor gene has been mapped to chromosome 5q. Patients with FAP develop numerous  
35 colon polyps at an early age, usually by late adolescence. Colon polyps are believed to be precursors

to adenocarcinoma of the colon. If left untreated, almost all carriers of the mutated gene will develop colorectal cancer. The mutated gene associated with this syndrome is known as APC (adenomatous polyposis coli). The second known inherited syndrome is hereditary nonpolyposis colorectal cancer (HNPCC), which is caused by mutations in mismatch repair genes.

5        Although hereditary colon cancer syndromes occur in a small percentage of the population and most colorectal cancers are considered sporadic, knowledge from studies of the hereditary syndromes can be generally applied. For instance, somatic mutations in APC occur in at least 80% of indiscriminate colon tumors. APC mutations are thought to be the initiating event in the disease. Other mutations occur subsequently. Approximately 50% of colorectal cancers contain activating  
10 mutations in ras, while 85% contain inactivating mutations in p53. Changes in these genes lead to gene expression changes in colon cancer. Less is understood about downstream targets of these mutations and the role they may play in cancer development and progression.

#### Parkinson's disease

      Parkinson's disease (PD) is a neurodegenerative disorder characterized by the progressive  
15 degeneration of the dopaminergic nigrostriatal pathway, and the presence of Lewy bodies. Genetic linkages to chromosomes 2p4, 4p5, and three loci on 1q6-8 have been identified (Gwinn-Hardy K. (2002) Mov. Disord. 17:645-656). Clinical disorders classified as parkinsonism include PD, dementia with Lewy bodies (DLB), progressive supranuclear palsy (PSP), and essential tremor. Several neurodegenerative diseases share pathogenic mechanisms involving tau or synuclein  
20 aggregation. These disorders include Alzheimer's disease, and Pick's disease as well as PD and progressive supranuclear palsy (Hardy, J. (2001) J. Alzheimers Dis. 3:109-116). Several genetically distinct forms of PD can be caused by mutations in single genes. Genes for monogenically inherited forms of Parkinson's disease have been mapped and/or cloned. In some families with autosomal dominant inheritance and typical Lewy-body pathology, mutations have been identified in the gene  
25 for alpha-synuclein. Aggregation of this protein in Lewy-bodies may be a crucial step in the molecular pathogenesis of familial and sporadic PD. On the other hand, mutations in the parkin gene cause early-onset autosomal recessive parkinsonism in which nigral degeneration is not accompanied by Lewy-body formation. Parkin-mutations appear to be a common cause of PD in patients with very early onset. Parkin has been implicated in the cellular protein degradation pathways, as it has been  
30 shown that it functions as a ubiquitin ligase. A mutation in the gene for ubiquitin C-terminal hydrolase L1 in this pathway has been identified in another small family with PD. Other loci have been mapped to chromosome 2p and 4p, respectively, in families with dominantly inherited PD. These early-onset forms differ from the common sporadic form of PD. It is widely believed that a combination of interacting genetic and environmental causes may be responsible in the majority of  
35 PD-cases (Gasser, T. (2001) J. Neurol. 2001 248:833-840).

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta.

5

### SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, proteins associated with cell growth, differentiation, and death, referred to collectively as 'CGDD' and individually as 'CGDD-1,' 'CGDD-2,' 'CGDD-3,' 'CGDD-4,' 'CGDD-5,' 'CGDD-6,' 'CGDD-7,' 'CGDD-8,'  
 10 'CGDD-9,' 'CGDD-10,' 'CGDD-11,' 'CGDD-12,' 'CGDD-13,' 'CGDD-14,' 'CGDD-15,' 'CGDD-16,' 'CGDD-17,' 'CGDD-18,' 'CGDD-19,' 'CGDD-20,' 'CGDD-21,' 'CGDD-22,' 'CGDD-23,' 'CGDD-24,' 'CGDD-25,' 'CGDD-26,' 'CGDD-27,' 'CGDD-28,' 'CGDD-29,' 'CGDD-30,' 'CGDD-31,' 'CGDD-32,' and 'CGDD-33' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions.  
 15 Embodiments also provide methods for utilizing the purified proteins associated with cell growth, differentiation, and death and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified proteins associated with cell growth, differentiation, and death and/or their encoding polynucleotides for investigating the pathogenesis of  
 20 diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID  
 25 NO:1-33, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-33.

30 Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, c) a biologically active fragment of a polypeptide  
 35 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, and d) an

immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-33. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:34-66.

5           Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group  
10 consisting of SEQ ID NO:1-33, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

15           Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, c) a biologically active fragment of a polypeptide having an amino  
20 acid sequence selected from the group consisting of SEQ ID NO:1-33, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering  
25 the polypeptide so expressed.

          Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an  
30 amino acid sequence selected from the group consisting of SEQ ID NO:1-33, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33.

          Still yet another embodiment provides an isolated polynucleotide selected from the group  
35 consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group

consisting of SEQ ID NO:34-66, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 5 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66, b) a 10 polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides 15 comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, 20 the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66, b) a 25 polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target 30 polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, b) a polypeptide comprising a 35 naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an



amino acid sequence selected from the group consisting of SEQ ID NO:1-33, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, and a pharmaceutically acceptable excipient.

- 5 In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-33. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional CGDD, comprising administering to a patient in need of such treatment the composition.

- Yet another embodiment provides a method for screening a compound for effectiveness as an  
10 agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ  
15 ID NO:1-33, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a  
20 disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment the composition.

- Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, b) a polypeptide  
25 comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33. The method comprises a)  
30 exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional CGDD, comprising administering to a patient in need of such treatment the composition.

- 35 Another embodiment provides a method of screening for a compound that specifically binds

to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, c) a biologically active  
5 fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the  
10 polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an  
15 amino acid sequence selected from the group consisting of SEQ ID NO:1-33, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the  
20 polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

25 Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target  
30 polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20  
35 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide

comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide

5 complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at

10 least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d)

15 comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

20 Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the

25 matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble

30 polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and

35 polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

### DESCRIPTION OF THE INVENTION

5 Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

10 As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same  
15 meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might  
20 be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

### DEFINITIONS

"CGDD" refers to the amino acid sequences of substantially purified CGDD obtained from  
25 any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CGDD. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CGDD either by directly interacting with  
30 CGDD or by acting on components of the biological pathway in which CGDD participates.

An "allelic variant" is an alternative form of the gene encoding CGDD. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to  
35 allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides.

Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding CGDD include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CGDD or a polypeptide with at least one functional characteristic of CGDD. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CGDD, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding CGDD. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CGDD. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CGDD is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms “amino acid” and “amino acid sequence” can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where “amino acid sequence” is recited to refer to a sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

“Amplification” relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of CGDD. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CGDD either by directly interacting with CGDD or by acting on components of the biological pathway in which CGDD participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CGDD polypeptides can be prepared using intact polypeptides or using

fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH<sub>2</sub>), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides

having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic CGDD, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding CGDD or fragments of CGDD may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys, Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
5	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
10	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
15	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
20	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the



evolution of new protein functions.

A "fragment" is a unique portion of CGDD or a polynucleotide encoding CGDD which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:34-66 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:34-66, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:34-66 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:34-66 from related polynucleotides. The precise length of a fragment of SEQ ID NO:34-66 and the region of SEQ ID NO:34-66 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-33 is encoded by a fragment of SEQ ID NO:34-66. A fragment of SEQ ID NO:1-33 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-33. For example, a fragment of SEQ ID NO:1-33 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-33. The precise length of a fragment of SEQ ID NO:1-33 and the region of SEQ ID NO:1-33 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical residue matches between at least two polynucleotide sequences aligned

using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more  
 5 computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in  
 10 Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic  
 15 Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2  
 20 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version  
 25 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

30 *Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example,  
 35 as defined by a particular SEQ ID number, or may be measured over a shorter length, for example,

over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to  
5 describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

10 The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the  
15 site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide  
20 sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap  
25 penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for  
30 example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

35 *Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present

invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents  
5 include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such  
10 similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters,  
15 chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune  
20 disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CGDD which is capable of eliciting an immune response when introduced into a living organism, for example, a  
25 mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of CGDD which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

30 The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of CGDD. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CGDD.

35 The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide,

polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an CGDD may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of CGDD.

"Probe" refers to nucleic acids encoding CGDD, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4<sup>th</sup> ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR

Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

5       Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the  
10   PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to  
15   avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing  
20   selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or  
25   partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

      A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more  
30   commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a  
35   vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A “regulatory element” refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

“Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An “RNA equivalent,” in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term “sample” is used in its broadest sense. A sample suspected of containing CGDD, nucleic acids encoding CGDD, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms “specific binding” and “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term “substantially purified” refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A “substitution” refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

“Substrate” refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells,



trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient  
5 cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term  
10 "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic  
15 acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The  
20 term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques  
25 for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-  
30 1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have  
35 significant identity to a reference molecule, but will generally have a greater or lesser number of

polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

## THE INVENTION

Various embodiments of the invention include new human proteins associated with cell growth, differentiation, and death (CGDD), the polynucleotides encoding CGDD, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ

ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the  
 5 probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding  
 10 Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Accelrys, Burlington MA). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column  
 15 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are proteins associated with cell growth, differentiation, and death. For example, SEQ ID NO:13 is 97% identical, from residue S220 to  
 20 residue A552 and 95% identical, from residue M1 to residue I240, to human DOC-2 (GenBank ID g1297330) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $2.5e-297$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:13 also has homology to proteins that are involved in cell proliferation and adhesion, and are DOC-2, as determined by BLAST analysis using  
 25 the PROTEOME database. SEQ ID NO:13 also contains a phosphotyrosine interaction (PTB/PID) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from other BLAST analyses provide further corroborative evidence that SEQ ID NO:13 is a DOC-2.

In another example, SEQ ID NO:19 is 100% identical, from residue M1 to residue C191, to  
 30 human D53 (GenBank ID g1469920) as determined by BLAST. The BLAST probability score is  $4.6e-92$ . SEQ ID NO:19 also has homology to proteins that play a role in cell proliferation, and are tumor protein D52-like 1 proteins, as determined by BLAST analysis using the PROTEOME database. Data from other BLAST analyses provide further corroborative evidence that SEQ ID NO:19 is a tumor protein D52 protein.

35 In yet another example, SEQ ID NO:26 is 93% identical, from residue I46 to residue I109,

and 100% identical, from residue M1 to residue E49, to murine mm-Mago (GenBank ID g2330011) as determined by BLAST. The BLAST probability score is  $4.6e-54$ . SEQ ID NO:26 also has homology to proteins that have developmental function, and are Mago-nashi proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:26 also contains a Mago-nashi domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. (See Table 3.) Data from additional BLAST analyses provide further corroborative evidence that SEQ ID NO:26 is a Mago-nashi protein.

In a further example, SEQ ID NO:32 is 51% identical, from residue E11 to residue Y682, to human cytoplasmic protein, Ropp120 (GenBank ID g14583268) as determined by BLAST. The BLAST probability score is  $3.6e-188$ . SEQ ID NO:32 also has homology to proteins that are localized to the cytoskeleton, and are similar to WD-repeat microtubule-associated proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:32 also contains seven G-beta repeat WD domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. (See Table 3.) Data from BLIMPS and other BLAST analyses provide further corroborative evidence that SEQ ID NO:32 is a human WD-repeat microtubule-associated protein. SEQ ID NO:1-12, SEQ ID NO:14-18, SEQ ID NO:20-25, SEQ ID NO:27-31, and SEQ ID NO:33 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-33 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:34-66 or that distinguish between SEQ ID NO:34-66 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from

the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as

FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3...</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as

FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the

GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to  
 5 construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte  
 10 identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid  
 15 encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses CGDD variants. Various embodiments of CGDD variants  
 20 can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the CGDD amino acid sequence, and can contain at least one functional or structural characteristic of CGDD.

Various embodiments also encompass polynucleotides which encode CGDD. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected  
 25 from the group consisting of SEQ ID NO:34-66, which encodes CGDD. The polynucleotide sequences of SEQ ID NO:34-66, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding CGDD. In particular,  
 30 such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding CGDD. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:34-66 which has at least about 70%, or  
 35 alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:34-66. Any one of the

polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of CGDD.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding CGDD. A splice variant may have portions which have significant  
5 sequence identity to a polynucleotide encoding CGDD, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding CGDD over its entire length; however, portions of the splice  
10 variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding CGDD. For example, a polynucleotide comprising a sequence of SEQ ID NO:42 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:43; a polynucleotide comprising a sequence of SEQ ID NO:45 is a splice variant of a polynucleotide comprising a sequence of SEQ ID  
15 NO:55; and a polynucleotide comprising a sequence of SEQ ID NO:56 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:57. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of CGDD.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the  
20 genetic code, a multitude of polynucleotide sequences encoding CGDD, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the  
25 polynucleotide sequence of naturally occurring CGDD, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode CGDD and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring CGDD under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding CGDD or its  
30 derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CGDD and its derivatives without altering the encoded amino acid sequences include the  
35 production of RNA transcripts having more desirable properties, such as a greater half-life, than

transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode CGDD and CGDD derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell  
5 systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding CGDD or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:34-66 and fragments thereof, under various conditions of stringency (Wahl, G.M. and  
10 S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment  
15 of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ  
20 Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and  
25 Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding CGDD may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic  
30 DNA within a cloning vector (Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known  
35 sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR



Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may  
5 use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of  
10 about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence  
15 into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the  
20 emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

25 In another embodiment of the invention, polynucleotides or fragments thereof which encode CGDD may be cloned in recombinant DNA molecules that direct expression of CGDD, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express CGDD.

30 The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter CGDD-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed  
35 mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation

patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. 5 Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CGDD, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired 10 properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of 15 homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding CGDD may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) 20 Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, CGDD itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated 25 synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of CGDD, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid 30 chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active CGDD, the polynucleotides encoding CGDD or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains 35 the necessary elements for transcriptional and translational control of the inserted coding sequence in

a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding CGDD. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding CGDD. Such signals

5 include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding CGDD and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should

10 be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression

15 vectors containing polynucleotides encoding CGDD and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express

20 polynucleotides encoding CGDD. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors

25 (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc.

30 Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-

242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding CGDD. For example, routine cloning, subcloning, and propagation of polynucleotides encoding CGDD can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding CGDD into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of CGDD are needed, e.g. for the production of antibodies, vectors which direct high level expression of CGDD may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CGDD. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of CGDD. Transcription of polynucleotides encoding CGDD may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding CGDD may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CGDD in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus

(RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are  
5 constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of CGDD in cell lines is preferred. For example, polynucleotides encoding CGDD can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or  
10 endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be  
15 propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*<sup>-</sup> and *apr*<sup>-</sup> cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or  
20 herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular  
25 requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$ -glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995)  
30 Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CGDD is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding CGDD can be identified by the absence of marker gene function.  
35 Alternatively, a marker gene can be placed in tandem with a sequence encoding CGDD under the

control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding CGDD and that express CGDD may be identified by a variety of procedures known to those of skill in the art. These procedures  
5 include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CGDD using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques  
10 include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CGDD is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect.  
15 IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled  
20 hybridization or PCR probes for detecting sequences related to polynucleotides encoding CGDD include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding CGDD, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA  
25 polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

30 Host cells transformed with polynucleotides encoding CGDD may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CGDD may be designed to contain signal sequences which  
35 direct secretion of CGDD through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a “prepro” or “pro” form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding CGDD may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CGDD protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CGDD activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CGDD encoding sequence and the heterologous protein sequence, so that CGDD may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled CGDD may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

CGDD, fragments of CGDD, or variants of CGDD may be used to screen for compounds that specifically bind to CGDD. One or more test compounds may be screened for specific binding to CGDD. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened

for specific binding to CGDD. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of CGDD can be used to screen for binding of test compounds, such as antibodies, to CGDD, a variant of CGDD, or a combination of CGDD and/or one or more variants CGDD. In an embodiment, a variant of CGDD can be used to screen for compounds that bind to a variant of CGDD, but not to CGDD having the exact sequence of a sequence of SEQ ID NO:1-33. CGDD variants used to perform such screening can have a range of about 50% to about 99% sequence identity to CGDD, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to CGDD can be closely related to the natural ligand of CGDD, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor CGDD (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to CGDD can be closely related to the natural receptor to which CGDD binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for CGDD which is capable of propagating a signal, or a decoy receptor for CGDD which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG<sub>1</sub> (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to CGDD, fragments of CGDD, or variants of CGDD. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of CGDD. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of CGDD. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of CGDD.

In an embodiment, anticalins can be screened for specific binding to CGDD, fragments of



CGDD, or variants of CGDD. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

10 In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit CGDD involves producing appropriate cells which express CGDD, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing CGDD or cell membrane fractions which contain CGDD are then contacted with a test compound and binding, stimulation, or inhibition of activity of either CGDD or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with CGDD, either in solution or affixed to a solid support, and detecting the binding of CGDD to the compound.

20 Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

35 CGDD, fragments of CGDD, or variants of CGDD may be used to screen for compounds that modulate the activity of CGDD. Such compounds may include agonists, antagonists, or partial or

inverse agonists. In one embodiment, an assay is performed under conditions permissive for CGDD activity, wherein CGDD is combined with at least one test compound, and the activity of CGDD in the presence of a test compound is compared with the activity of CGDD in the absence of the test compound. A change in the activity of CGDD in the presence of the test compound is indicative of a compound that modulates the activity of CGDD. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising CGDD under conditions suitable for CGDD activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CGDD may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding CGDD or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding CGDD may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding CGDD can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding CGDD is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and

treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress CGDD, e.g., by secreting CGDD in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

### THERAPEUTICS

5 Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CGDD and proteins associated with cell growth, differentiation, and death. In addition, examples of tissues expressing CGDD can be found in Table 6 and can also be found in Example XI. Therefore, CGDD appears to play a role in cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta. In the treatment of disorders associated with increased CGDD expression or activity, it is desirable to decrease the expression or activity of CGDD. In the treatment of disorders associated with decreased CGDD expression or activity, it is desirable to increase the expression or activity of CGDD.

Therefore, in one embodiment, CGDD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess,

suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis,

5 encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental

10 disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,

15 allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's

20 syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome,

25 complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune

30 disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puberty,

35 retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaeochromocytomas,

paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumors; and a disorder of the placenta such as preeclampsia, choriocarcinoma, abruptio placentae, placenta previa, placental or maternal floor infarction, placenta accreta, increta, and percreta, extrachorial placentas, chorangioma, chorangiosis, chronic villitis, placental villous edema, widespread fibrosis of the terminal villi, intervillous thrombi, hemorrhagic endovasculitis, erythroblastosis fetalis, and nonimmune fetal hydrops.

In another embodiment, a vector capable of expressing CGDD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified CGDD in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CGDD may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CGDD including, but not limited to, those listed above.

In a further embodiment, an antagonist of CGDD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CGDD. Examples of such disorders include, but are not limited to, those cell proliferative disorders including cancer, developmental disorders, neurological disorders, autoimmune/inflammatory disorders, reproductive disorders, and disorders of the placenta described above. In one aspect, an antibody which specifically binds CGDD may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CGDD.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CGDD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CGDD including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CGDD may be produced using methods which are generally known in the art. In particular, purified CGDD may be used to produce antibodies or to screen libraries of

pharmaceutical agents to identify those which specifically bind CGDD. Antibodies to CGDD may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e.,  
5 those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels,  
10 dromedaries, llamas, humans, and others may be immunized by injection with CGDD or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol.  
15 Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CGDD have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or  
20 fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of CGDD amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CGDD may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not  
25 limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the  
30 splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CGDD-specific single chain  
35 antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be

generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for CGDD may also be generated. For example, such fragments include, but are not limited to,  $F(ab)_2$  fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the  $F(ab)_2$  fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CGDD and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CGDD epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CGDD. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of CGDD-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CGDD epitopes, represents the average affinity, or avidity, of the antibodies for CGDD. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular CGDD epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the CGDD-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CGDD, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to

determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CGDD-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and  
5 guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding CGDD, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA,  
10 RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding CGDD. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CGDD (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense  
15 sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102:469-475; Scanlon, K.J. et al. (1995) 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as  
20 retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) *Blood* 76:271; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) *Br. Med. Bull.* 51:217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87:1308-1315; Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25:2730-2736).

25 In another embodiment of the invention, polynucleotides encoding CGDD may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency  
30 (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii)  
35 express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated



cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides* 5 *brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in CGDD expression or regulation causes disease, the expression of CGDD from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in 10 CGDD are treated by constructing mammalian expression vectors encoding CGDD and introducing these vectors by mechanical means into CGDD-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev.* 15 *Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. R  capon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of CGDD include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), 20 and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CGDD may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and 25 H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding CGDD from a normal individual.

30 Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. 35 (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of

these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to CGDD expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding CGDD under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding CGDD to cells which have one or more genetic abnormalities with respect to the expression of CGDD. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CGDD to target cells which have one or more genetic abnormalities with respect to the expression of CGDD. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CGDD to cells of the central nervous system, for which HSV has a

tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S.

- 5 Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22.
- 10 For HSV vectors, see also Goins, W.F. et al. (1999; *J. Virol.* 73:519-532) and Xu, H. et al. (1994; *Dev. Biol.* 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

- 15 In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding CGDD to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid
- 20 proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for CGDD into the alphavirus genome in place of the capsid-coding region results in the production of a large number of CGDD-coding RNAs and the synthesis of high levels of CGDD in vector transduced cells. While
- 25 alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of CGDD into a variety of cell types. The specific transduction of a subset of
- 30 cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

- Oligonucleotides derived from the transcription initiation site, e.g., between about positions
- 35 -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition

can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and  
5 Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme  
10 molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding CGDD.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA,  
15 GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

20 Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding CGDD. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA  
25 polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase  
30 linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

35 In other embodiments of the invention, the expression of one or more selected

polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene  
5 bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods  
10 described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. SiRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. SiRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs  
15 appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

SiRNA can either be generated indirectly by introduction of dsRNA into the targeted cell, or directly by mammalian transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory  
20 methods). Suitable SiRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target  
25 siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target  
30 sequences with significant homology to other coding sequences can be eliminated from consideration. The selected SiRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished  
35 using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods

known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the  
5 target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene, can be determined by northern analysis methods using, for example, the  
10 NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a  
15 compound which is effective in altering expression of a polynucleotide encoding CGDD. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide  
20 sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased CGDD expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding CGDD may be therapeutically useful, and in the treatment of disorders associated with decreased CGDD expression or activity, a compound which specifically promotes  
25 expression of the polynucleotide encoding CGDD may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary  
30 library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CGDD is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted  
35 biochemical system. Alterations in the expression of a polynucleotide encoding CGDD are assayed

by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding CGDD. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of CGDD, antibodies to CGDD, and mimetics, agonists, antagonists, or inhibitors of CGDD.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form.

These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung  
5 have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination  
10 of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising CGDD or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CGDD or a fragment thereof may be joined to a short cationic N-  
15 terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys,  
20 or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CGDD or fragments thereof, antibodies of CGDD, and agonists, antagonists or inhibitors of CGDD, which  
25 ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large  
30 therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

35 The exact dosage will be determined by the practitioner, in light of factors related to the



subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

### DIAGNOSTICS

In another embodiment, antibodies which specifically bind CGDD may be used for the diagnosis of disorders characterized by expression of CGDD, or in assays to monitor patients being treated with CGDD or agonists, antagonists, or inhibitors of CGDD. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CGDD include methods which utilize the antibody and a label to detect CGDD in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CGDD, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CGDD expression. Normal or standard values for CGDD expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to CGDD under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CGDD expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding CGDD may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CGDD may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CGDD, and to monitor regulation of CGDD levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding CGDD or closely related molecules may be used to identify nucleic acid sequences which encode CGDD. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CGDD, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CGDD encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:34-66 or from genomic sequences including promoters, enhancers, and introns of the CGDD gene.

Means for producing specific hybridization probes for polynucleotides encoding CGDD include the cloning of polynucleotides encoding CGDD or CGDD derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding CGDD may be used for the diagnosis of disorders associated with expression of CGDD. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal

disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system

5 disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic

10 nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses,

15 postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal

20 dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation,

25 osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a reproductive disorder such as a disorder

30 of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer

35 of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of

the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puberty, retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumors; and a disorder of the placenta such as preeclampsia, choriocarcinoma, abruptio placentae, placenta previa, placental or maternal floor infarction, placenta accreta, increta, and percreta, extrachorial placentas, chorangioma, chorangiosis, chronic villitis, placental villous endema, widespread fibrosis of the terminal villi, intervillous thrombi, hemorrhagic endovascularitis, erythroblastosis fetalis, and nonimmune fetal hydrops. Polynucleotides encoding CGDD may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CGDD expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding CGDD may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding CGDD may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding CGDD in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CGDD, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CGDD, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several

days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CGDD may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding CGDD, or a fragment of a polynucleotide complementary to the polynucleotide encoding CGDD, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding CGDD may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding CGDD are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis,

sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of CGDD include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, CGDD, fragments of CGDD, or antibodies specific for CGDD may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by

quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of  
5 transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines,  
10 biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental  
15 compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or  
20 signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with  
25 different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>). Therefore, it is important and  
30 desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present  
35 invention may be quantified. The transcript levels in the treated biological sample are compared with

levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for CGDD to quantify the levels of CGDD expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which



alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding CGDD may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154).

Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

5       Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CGDD on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of  
10 DNA associated with that disorder and thus may further positional cloning efforts.

*In situ* hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is  
15 valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant  
20 invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CGDD, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a  
25 solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CGDD and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a  
30 solid substrate. The test compounds are reacted with CGDD, or fragments thereof, and washed. Bound CGDD is then detected by methods well known in the art. Purified CGDD can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing  
35 antibodies capable of binding CGDD specifically compete with a test compound for binding CGDD.

In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CGDD.

In additional embodiments, the nucleotide sequences which encode CGDD may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/340,747, U.S. Ser. No. 60/342,761, U.S. Ser. No. 60/349,705, U.S. Ser. No. 60/354,764, and U.S. Ser. No. 60/356,216, are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or

enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-  
5 TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Invitrogen.

## 10 II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid,  
15 QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal  
20 cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

25 Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared  
30 using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI  
35 protocols and base calling software; or other sequence analysis systems known in the art. Reading

frames within the cDNA sequences were identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing  
 5 vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*,  
 10 *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002)  
 15 Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences,  
 20 or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of  
 25 the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences  
 30 are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

35 Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of

Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:34-66. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

#### **IV. Identification and Editing of Coding Sequences from Genomic DNA**

Putative proteins associated with cell growth, differentiation, and death were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpr and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode proteins associated with cell growth, differentiation, and death, the encoded polypeptides were analyzed by querying against PFAM models for proteins associated with cell growth, differentiation, and death. Potential proteins associated with cell growth, differentiation, and death were also identified by homology to Incyte cDNA sequences that had been annotated as proteins associated with cell growth, differentiation, and death. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted

coding sequences.

## V. Assembly of Genomic Sequence Data with cDNA Sequence Data

### "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene  
5 identification program described in Example IV. Partial cDNAs assembled as described in Example  
III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan  
exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm  
based on graph theory and dynamic programming to integrate cDNA and genomic information,  
generating possible splice variants that were subsequently confirmed, edited, or extended to create a  
10 full length sequence. Sequence intervals in which the entire length of the interval was present on  
more than one sequence in the cluster were identified, and intervals thus identified were considered to  
be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic  
sequences, then all three intervals were considered to be equivalent. This process allows unrelated  
but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals  
15 thus identified were then "stitched" together by the stitching algorithm in the order that they appear  
along their parent sequences to generate the longest possible sequence, as well as sequence variants.  
Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or  
genomic sequence to genomic sequence) were given preference over linkages which change parent  
type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared  
20 by BLAST analysis to the genpept and gbprl public databases. Incorrect exons predicted by Genscan  
were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended  
with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST  
25 analysis. First, partial cDNAs assembled as described in Example III were queried against public  
databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases  
using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST  
analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in  
Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs  
30 (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions  
may occur in the chimeric protein with respect to the original GenBank protein homolog. The  
GenBank protein homolog, the chimeric protein, or both were used as probes to search for  
homologous genomic sequences from the public human genome databases. Partial DNA sequences  
were therefore "stretched" or extended by the addition of homologous genomic sequences. The  
35 resultant stretched sequences were examined to determine whether it contained a complete gene.

## VI. Chromosomal Mapping of CGDD Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:34-66 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched  
 5 SEQ ID NO:34-66 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment  
 10 of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in  
 15 humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified  
 20 disease genes map within or in proximity to the intervals indicated above.

### Association of CGDD Polynucleotides with Parkinson's Disease

Several genes have been identified as showing linkage to autosomal dominant forms of Parkinson's Disease (PD). PD is a common neurodegenerative disorder causing bradykinesia, resting tremor, muscular rigidity, and postural instability. Cytoplasmic eosinophilic inclusions  
 25 called Lewy bodies, and neuronal loss especially in the substantia nigra pars compacta, are pathological hallmarks of PD (Valente, E.M. et al (2001) Am. J. Hum. Genet. 68:895-900). Lewy body Parkinson disease has been thought to be a specific autosomal dominant disorder (Wakabayashi, K. et al. (1998) Acta Neuropath. 96:207-210). Juvenile parkinsonism may be a specific autosomal recessive disorder (Matsumine, H. et al. (1997) Am. J. Hum. Genet. 60:  
 30 588-596, 1997). (Online Mendelian Inheritance in Man, OMIM. Johns Hopkins University, Baltimore, MD. MIM Number: 168600: Sept. 9, 2002: . World Wide Web URL: <http://www.ncbi.nlm.nih.gov/omim/>)

Association of a disease with a chromosomal locus can be determined by lod score. Lod score is a statistical method used to test the linkage of two or more loci within families having a  
 35 genetic disease. The lod score is the logarithm to base 10 of the odds in favor of linkage.



Linkage is defined as the tendency of two genes located on the same chromosome to be inherited together through meiosis (*Genetics in Medicine*, Fifth Edition, (1991) Thompson, M.W. et al., W.B. Saunders Co. Philadelphia). A lod score of +3 or greater (1000:1 odds in favor of linkage) indicates a probability of 1 in 1000 that a particular marker was found solely by chance in affected  
5 individuals, which is strong evidence that two genetic loci are linked.

One such gene implicated in PD is PARK3, which maps to 2p13 (Gasser, T. et al. (1998) *Nature Genet.* 18:262-265). A marker at chromosomal position D2S441 was found to have a lod score of 3.2 in the region of PARK3. This marker supported the disease association of PARK3 in the chromosomal interval from D2S134 to D2S286 (Gasser et al., *supra*). Markers located  
10 within chromosomal intervals D2S134 and D2S286, which map between 83.88 to 94.05 centiMorgans on the short arm of chromosome 2, were used to identify genes that map in the region between D2S134 and D2S286.

A second PD gene, implicated in early-onset recessive parkinsonism, is PARK6, located on chromosome 1 at 1p35-1p36. Several markers were obtained with lod scores greater than 3  
15 including D1S199, D1S2732, D1S2828, D1S478, D1S2702, D1S2734, D1S2674 (Valente, E.M. et al, *supra*). These markers were used to determine the PD-relevant range of chromosome loci and identify sequences that map to chromosome 1 between D1S199 and D1S2885.

Restriction fragment length polymorphism (RFLP) markers shown to be near regions of DNA known as sequence-tagged sites (STS), have been mapped to NT\_Contigs generated by the Human  
20 Genome Project using ePCR (Schuler, G.D. (1997) *Genome Research* 7: 541-550, and (1998) *Trends Biotechnol.* 16(11):456-9). Contigs containing regions of DNA with known disease-associated markers are therefore used to identify CGDD sequences that map to disease-associated regions of the genome.

Polynucleotides encoding CGDD were mapped to NT\_Contigs. Contigs longer than 1Mb  
25 were broken into subcontigs of 1Mb length with overlapping sections of 100kb. A preliminary step used an algorithm, similar to MEGABLAST, to define the mRNA sequence /masked genomic DNA contig pairings. The cDNA/genomic pairings identified by the first algorithm were confirmed, and the CGDD polynucleotides mapped to DNA contigs, using SIM4 (Florea, L. et al. (1998) *Genome Res.* 8:967-74, version May 2000) which had been optimized for high throughput processing and  
30 strand assignment confidence). The SIM4 output of the mRNA sequence/genomic contig pairs was further processed to determine the correct location of the CGDD polynucleotides on the genomic contig, as well as their strand identity.

Both SEQ ID NO:38 and SEQ ID NO:48 were mapped to NT\_Contig NT\_004782 from Genbank release February, 2002, covering a 14.87 Mb region of the genome that also contains PD-  
35 associated genetic markers D1S199 and D1S2885. The maximum distance between both SEQ ID

NO:38 and SEQ ID NO:48 and markers D1S199 and D1S2885, therefore, is 14.87 Mb. Thus, both SEQ ID NO:38 and SEQ ID NO:48 are in proximity with genetic markers shown to consistently associate with PD.

## VII. Analysis of Polynucleotide Expression

5 Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

10 Analogous computer techniques applying BLAST were used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$15 \quad \frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is  
 20 calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate  
 25 the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79%  
 30 identity and 100% overlap.

Alternatively, polynucleotides encoding CGDD are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into  
 35 one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive

system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract.

The number of libraries in each category is counted and divided by the total number of libraries

- 5 across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding CGDD. cDNA sequences and cDNA library/tissue  
10 information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

### VIII. Extension of CGDD Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to  
15 initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

20 Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ ,  
25 and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2:  
30 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar,  
35 Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II

(Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

5 The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were  
10 religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

15 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries  
20 were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used  
25 to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### **IX. Identification of Single Nucleotide Polymorphisms in CGDD Encoding Polynucleotides**

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:34-66 using the LIFESEQ database (Incyte Genomics). Sequences from the  
30 same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice  
35 variants. An automated procedure of advanced chromosome analysis analysed the original

chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to  
5 contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46  
10 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele  
15 frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

#### **X. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:34-66 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base  
20 pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25  
25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon  
30 membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### **35 XI. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### **Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte Genomics). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (Clontech, Palo Alto CA) and after

combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### **Microarray Preparation**

5            Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

10           Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C  
15    oven.

            Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

20           Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### **Hybridization**

            Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly  
30    larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

#### **Detection**

35           Reporter-labeled hybridization complexes are detected with a microscope equipped with an

Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte Genomics). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be differentially expressed.



**Expression**

For example, SEQ ID NO:38 was significantly under-expressed in the breast tumor cell lines when compared to a normal breast cell line. The gene expression profile of a nonmalignant mammary epithelial cell line was compared to the gene expression profiles of breast carcinoma lines at different stages of tumor progression. Cell lines compared included: a) BT-20, a breast carcinoma cell line derived in vitro from the cells emigrating out of thin slices of tumor mass isolated from a 74-year-old female, b) BT-474, a breast ductal carcinoma cell line that was isolated from a solid, invasive ductal carcinoma of the breast obtained from a 60-year-old woman, c) BT-483, a breast ductal carcinoma cell line that was isolated from a papillary invasive ductal tumor obtained from a 23-year-old normal, menstruating, parous female with a family history of breast cancer, d) Hs 578T, a breast ductal carcinoma cell line isolated from a 74-year-old female with breast carcinoma, e) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, f) MCF-10A, a breast mammary gland (luminal ductal characteristics) cell line isolated from a 36-year-old woman with fibrocystic breast disease, g) MDA-MB-468, a breast adenocarcinoma cell line isolated from the pleural effusion of a 51-year-old female with metastatic adenocarcinoma of the breast, and h) HMEC, a primary breast epithelial cell line isolated from a normal donor. For example, SEQ ID NO:38 was decreased at least 2.3-fold in five (BT-474, BT-483, MCF-10A, MCF7, and BT-20) of the seven cell lines tested (described above) when compared with HMEC cells whether grown in medium with or without growth factors and nutrients. Therefore, in various embodiments, SEQ ID NO:38 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

As with most tumors, prostate cancer develops through a multistage progression ultimately resulting in an aggressive tumor phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells. Androgen responsive cells become hyperplastic and evolve into early-stage tumors. Although early-stage tumors are often androgen sensitive and respond to androgen ablation, a population of androgen independent cells evolve from the hyperplastic population. These cells represent a more advanced form of prostate tumor that may become invasive and potentially become metastatic to the bone, brain, or lung. PrEC is a primary prostate epithelial cell line isolated from a normal donor. DU 145 is a prostate carcinoma cell line isolated from a metastatic site in the brain of 69-year old male with widespread metastatic prostate carcinoma. DU 145 has no detectable sensitivity to hormones; forms colonies in semi-solid medium; is only weakly positive for acid phosphatase; and cells are negative for prostate specific antigen (PSA). LNCaP is a prostate carcinoma cell line isolated from a lymph node biopsy of a 50-year-old male with metastatic prostate carcinoma. LNCaP cells express prostate specific antigens, produce

prostatic acid phosphatase, and express androgen receptors. PC-3 is a prostate adenocarcinoma cell line that was isolated from a metastatic site in the bone of a 62- year-old male with grade IV prostate adenocarcinoma. Starved: cells were grown in basal media in the absence of growth factors and hormones. In a further example, expression of Seq ID NO:38 was decreased at least two-fold in two  
5 (DU145 and LNCaP) of the three cell lines tested (described above) when compared with PrEC cells. Therefore, in various embodiments, SEQ ID NO:38 can be used for one or more of the following: i) monitoring treatment of prostate cancer, ii) diagnostic assays for prostate cancer, and iii) developing therapeutics and/or other treatments for prostate cancer.

#### Preparation of Monocytes and Immature Dendritic Cells

10 Peripheral blood mononuclear cells (PBMCs) were isolated from freshly obtained peripheral blood of four healthy donors by centrifugation of the lymphocyte enriched blood fraction over a HYPAAQUE ficoll gradient (Sigma-Aldrich). The PBMCs were allowed to adhere to plastic in Iscove's Modified Dulbecco's Medium supplemented with 10% fetal bovine serum, 200 nM glutamine, and 200 nM each penicillin and streptomycin for 4 hours to separate monocytes from other  
15 nonadherent cells. One-half of the monocytes was cultured for seven days in 10 ng/ml GM-CSF (Peprotech Inc., Rocky Hill NJ), and the other half was cultured for seven days with 10 ng/ml GM-CSF and 10 ng/ml IL-4 (Peprotech Inc.) to produce immature dendritic cells.

#### Preparation of Mature Dendritic Cells

Monocytes were isolated as described above and were incubated with 10 ng/ml GM-CSF and  
20 10 ng/ml IL-4 for 13 days to generate immature DC. The DC were activated with anti-CD40 (Biodesign International, Kennebunk ME) at 10 ng/ml for 24 hours. The DC were divided into half, and each half further divided into thirds. Each third was cultured with soluble human TRANCE protein (Peprotech Inc.) at 10 ng/ml for 2, 8, and 24 hours, respectively, or cultured without TRANCE for 2, 8, and 24 hours, respectively. In yet a further example, SEQ ID NO:38 was increased at least  
25 8.5-fold in dendritic cells. Therefore, in various embodiments, SEQ ID NO:38 can be used for one or more of the following: i) monitoring treatment of immune disorders and related diseases and conditions, ii) diagnostic assays for immune disorders and related diseases and conditions, and iii) developing therapeutics and/or other treatments for immune disorders and related diseases and conditions.

30 In another example, the expression of SEQ ID NO:40 was increased at least four-fold in cancerous lung tissue compared to normal tissue from the same donor. Lung squamous cell carcinoma tissue was obtained from a 68-year-old female and matched with normal lung tissue obtained from grossly uninvolved tissue from the same donor. Therefore, in various embodiments, SEQ ID NO:40 can be used for one or more of the following: i) monitoring treatment of lung  
35 squamous cell carcinoma, ii) diagnostic assays for lung squamous cell carcinoma, and iii) developing

therapeutics and/or other treatments for lung squamous cell carcinoma.

In another example, the expression of SEQ ID NO:42 was decreased at least two-fold in cancerous lung tissue compared to normal tissue from the same donor. Lung squamous cell carcinoma tissue was obtained from a 75-year-old female and matched with normal lung tissue  
5 obtained from grossly uninvolved tissue from the same donor. Therefore, in various embodiments, SEQ ID NO:42 can be used for one or more of the following: i) monitoring treatment of lung squamous cell carcinoma, ii) diagnostic assays for lung squamous cell carcinoma, and iii) developing therapeutics and/or other treatments for lung squamous cell carcinoma. Matched normal and tumorigenic lung tissue samples for SEQ ID NO:40 and SEQ ID NO:42 are provided by the Roy  
10 Castle International Centre for Lung Cancer Research, Liverpool UK).

In another example, Jurkat is an acute T cell leukemia cell line that grows actively in the absence of external stimuli. Jurkat has been extensively used to study signaling in human T cells. PMA is a broad activator of the protein kinase C-dependent pathways. Ionomycin is a calcium ionophore that permits the entry of calcium in the cell, hence increasing the cytosolic calcium  
15 concentration. The combination of PMA and ionomycin activates two of the major signaling pathways used by mammalian cells to interact with their environment. In T cells, the combination of PMA and ionomycin mimics the type of secondary signaling events elicited during optimal B cell activation. The expression of SEQ ID NO:42 was downregulated at least two-fold in the Jurkat T-cell leukemia cell line that had been stimulated for one hour with 1 $\mu$ M PMA (phorbol 12-myristate  
20 13-acetate) and with ionomycin concentrations varying between 50 ng/ml and 10  $\mu$ g/ml when compared to untreated Jurkat cells in the absence of stimuli. Therefore, in various embodiments, SEQ ID NO:42 can be used for one or more of the following: i) monitoring treatment of T cell leukemia, ii) diagnostic assays for T cell leukemia and in signaling assays in human T cells, and iii) developing therapeutics and/or other treatments for T cell leukemia.

25 In a further example, SEQ ID NO:42 was decreased at least two-fold in one of seven cell lines (HMEC, MCF7, MDA-mb-231, Sk-BR-3, MDA-mb-435S, BT-20, and T-47D) when compared with MCF10A cells (See descriptions above.). Therefore, in various embodiments, SEQ ID NO:42 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

30 In another example, expression of SEQ ID NO:47 was decreased at least two-fold in four out of five squamous cell carcinoma tissues and in one lung adenocarcinoma as determined by microarray analysis. Normal lung tissue from the donors was compared to lung tumor from the same donor. Donors are as follows: lung squamous cell carcinoma tissue was obtained from a 68-year-old female, a 66-year-old male, and two different 73-year-old males; lung adenocarcinoma tissue was obtained  
35 from a 66-year-old female. Matched normal and tumorigenic lung tissue samples for SEQ ID NO:47

are provided by the Roy Castle International Centre for Lung Cancer Research, Liverpool UK). Therefore, in various embodiments, SEQ ID NO:47 can be used for one or more of the following: i) monitoring treatment of lung cancer and other cell proliferative disorders, ii) diagnostic assays for lung cancer and other cell proliferative disorders, and iii) developing therapeutics and/or other treatments for lung cancer and other cell proliferative disorders.

Human peripheral blood mononuclear cells (PBMC) can be classified into discrete cellular populations representing the major cellular components of the immune system. PBMCs contain about 52% lymphocytes (12% B lymphocytes, 40 % T lymphocytes {25 % CD4+ and 15 % CD8+}), 20% NK cells, 25% monocytes, and 3% various other cells including dendritic cells and progenitor cells. The proportions, as well as the biology of these cellular components tend to slightly vary between healthy individuals, depending on factors such as age, gender, past medical history, and genetic background. PBMC from the blood of 6 healthy volunteer donors were incubated for 24 hours in the presence of graded doses of beclomethasone dissolved in DMSO or in DMSO alone. The treated PBMC were compared to matching untreated PBMC maintained in culture for the same duration. In a further example, expression of SEQ ID NO:48 was down regulated at least two-fold in PBMC cells treated with doses equal to or higher than 1  $\mu$ M (1  $\mu$ M, 5  $\mu$ M, and 25  $\mu$ M) of beclomethasone as determined by microarray analysis. Therefore, in various embodiments, SEQ ID NO:48 can be used for one or more of the following: i) monitoring treatment of immune disorders and related diseases and conditions, ii) diagnostic assays for immune disorders and related diseases and conditions, and iii) developing therapeutics and/or other treatments for immune disorders and related diseases and conditions.

SEQ ID NO:58 was differentially expressed in human peripheral blood mononuclear cells (PBMCs) treated with SEB (Staphylococcal enterotoxin B), when compared to untreated PBMC cells. PBMCs can be classified into discrete cellular populations representing the major cellular components of the immune system. PBMCs contain about 52% lymphocytes (12% B lymphocytes, 40% T lymphocytes {25% CD4+ and 15% CD8+}), 20% NK cells, 25% monocytes, and 3% various cells that include dendritic cells and progenitor cells. The proportions, as well as the biology of these cellular components tend to vary slightly between healthy individuals, depending on factors such as age, gender, past medical history, and genetic background.

Staphylococcal exotoxins specifically activate human T cells, expressing an appropriate TCR-V $\beta$  chain. Although polyclonal in nature, T cells activated by Staphylococcal exotoxins require antigen presenting cells (APCs) to present the exotoxin molecules to the T cells and deliver the costimulatory signals required for optimum T cell activation. Although, Staphylococcal exotoxins must be presented to T cells by APCs, these molecules are not required to be processed by APC. Indeed, Staphylococcal exotoxins directly bind to a non-polymorphic portion of the human MHC

class II molecules, bypassing the need to capture, cleavage, and binding of the peptides to the polymorphic antigenic groove of the MHC class II molecules.

To evaluate the variation in gene expression in the PBMCs from healthy donors in response to treatment with SEB, the PBMCs from healthy volunteer donors were stimulated *in vitro* with SEB.

5 SEQ ID NO:58 was overexpressed by at least two-fold in PBMC cells treated with SEB for 24 hours and for 72 hours. SEQ ID NO:58 was overexpressed by at least two-fold in PBMC cells treated with SEB for 72 hours. Therefore, in various embodiments, SEQ ID NO:58 can be used for one or more of the following: i) monitoring treatment of immune disorders and related diseases and conditions, ii) diagnostic assays for immune disorders and related diseases and conditions, and iii) developing  
10 therapeutics and/or other treatments for immune disorders and related diseases and conditions.

SEQ ID NO:58 was differentially expressed in human colon tissue compared to normal colon tissue. While soft tissue sarcomas are relatively rare, more than 50% of new patients diagnosed with the disease will die from it. The molecular pathways leading to the development of sarcomas are relatively unknown, due to the rarity of the disease and variation in pathology. It is likely that  
15 numerous gene expression differences exist between sarcomas and normal tissue. To begin to delineate the pathways that might lead to sarcoma formation, differences in gene expression were examined between a colon tumor derived from a gastric sarcoma and normal colon tissue from the same donor. SEQ ID NO:58 was overexpressed by at least two-fold in colon tumor tissue as compared to normal colon tissue from the same donor. Therefore, in various embodiments, SEQ ID  
20 NO:58 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

SEQ ID NO:53 and SEQ ID NO:58 were differentially expressed in human lung tissue compared to normal lung tissue. Lung cancers are divided into four histopathologically distinct  
25 groups. Three groups (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) are classified as non-small cell lung cancers (NSCLCs). The fourth group of cancers is referred to as small cell lung cancer (SCLC). Collectively, NSCLCs account for approximately 70% of cases while SCLCs account for approximately 18% of cases. The molecular and cellular biology underlying the development and progression of lung cancer are incompletely understood. Deletions on  
30 chromosome3 are common in this disease and are thought to indicate the presence of a tumor suppressor gene in this region. Activating mutations in K-ras are commonly found in lung cancer and are the basis of one of the mouse models for the disease. Analysis of gene expression patterns associated with the development and progression of the disease will yield tremendous insight into the biology underlying this disease. Further, SEQ ID NO:53 and SEQ ID NO:58 were overexpressed by  
35 at least two-fold in two pair comparison experiments in which tissue from squamous cell carcinoma

of the lung was compared to uninvolved lung tissue. from the same donor. Therefore, in various embodiments, SEQ ID NO:53 and SEQ ID NO:58 can be used for one or more of the following: i) monitoring treatment of lung cancer, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

5 In a further example, the expression of SEQ ID NO:64 was increased at least two-fold in cancerous colon tissue compared to normal tissue from the same donor. Colon tissue was obtained from a 38-year-old male with invasive, poorly differentiated adenocarcinoma with metastases to two out of 13 lymph nodes surveyed, and matched with normal colon tissue obtained from grossly uninvolved colon tissue from the same donor. Matched normal and tumorigenic colon tissue samples  
10 are provided by the Huntsman Cancer Institute (Salt Lake City, UT). Therefore, in various embodiments, SEQ ID NO:64 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

## **XII. Complementary Polynucleotides**

15 Sequences complementary to the CGDD-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CGDD. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CGDD. To  
20 inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CGDD-encoding transcript.

## **XIII. Expression of CGDD**

25 Expression and purification of CGDD is achieved using bacterial or virus-based expression systems. For expression of CGDD in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory  
30 element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CGDD upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CGDD in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is  
35 replaced with cDNA encoding CGDD by either homologous recombination or bacterial-mediated

transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (SF9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, CGDD is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from CGDD at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified CGDD obtained by these methods can be used directly in the assays shown in Examples XVII and XVIII, where applicable.

#### XIV. Functional Assays

CGDD function is assessed by expressing the sequences encoding CGDD at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side

light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are  
5 discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of CGDD on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CGDD and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using  
10 magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CGDD and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### **XV. Production of CGDD Specific Antibodies**

CGDD substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g.,  
15 Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the CGDD amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is  
20 synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-  
25 Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CGDD activity by, for example, binding the peptide or CGDD to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **30 XVI. Purification of Naturally Occurring CGDD Using Specific Antibodies**

Naturally occurring or recombinant CGDD is substantially purified by immunoaffinity chromatography using antibodies specific for CGDD. An immunoaffinity column is constructed by covalently coupling anti-CGDD antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and  
35 washed according to the manufacturer's instructions.



Media containing CGDD are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CGDD (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CGDD binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CGDD is collected.

#### **XVII. Identification of Molecules Which Interact with CGDD**

CGDD, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CGDD, washed, and any wells with labeled CGDD complex are assayed. Data obtained using different concentrations of CGDD are used to calculate values for the number, affinity, and association of CGDD with the candidate molecules.

Alternatively, molecules interacting with CGDD are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; *Nature* 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

CGDD may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### **XVIII. Demonstration of CGDD Activity**

CGDD activity is demonstrated by measuring the induction of terminal differentiation or cell cycle progression when CGDD is expressed at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies, Gaithersburg, MD) and PCR 3.1 (Invitrogen, Carlsbad, CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu\text{g}$  of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2  $\mu\text{g}$  of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP) (Clontech, Palo Alto, CA), CD64, or a CD64-GFP fusion protein. Flow cytometry detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell cycle progression or terminal differentiation. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward

light scatter and 90 degree side light scatter; up or down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York, NY.

Alternatively, an *in vitro* assay for CGDD activity measures the transformation of normal human fibroblast cells overexpressing antisense CGDD RNA (Garkavtsev, I. and K. Riabowol (1997) Mol. Cell Biol. 17:2014-2019). cDNA encoding CGDD is subcloned into the pLNCX retroviral vector to enable expression of antisense CGDD RNA. The resulting construct is transfected into the ecotropic BOSC23 virus-packaging cell line. Virus contained in the BOSC23 culture supernatant is used to infect the amphotropic CAK8 virus-packaging cell line. Virus contained in the CAK8 culture supernatant is used to infect normal human fibroblast (Hs68) cells. Infected cells are assessed for the following quantifiable properties characteristic of transformed cells: growth in culture to high density associated with loss of contact inhibition, growth in suspension or in soft agar, formation of colonies or foci, lowered serum requirements, and ability to induce tumors when injected into immunodeficient mice. The activity of CGDD is proportional to the extent of transformation of Hs68 cells.

Alternatively, CGDD can be expressed in a mammalian cell line by transforming the cells with a eukaryotic expression vector encoding CGDD. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. To assay the cellular localization of CGDD, cells are fractionated as described by Jiang, H.P. et al. (1992; Proc. Natl. Acad. Sci. 89:7856-7860). Briefly, cells pelleted by low-speed centrifugation are resuspended in buffer (10 mM TRIS-HCl, pH 7.4/ 10 mM NaCl/ 3 mM MgCl<sub>2</sub>/ 5 mM EDTA with 10 ug/ml aprotinin, 10 ug/ml leupeptin, 10 ug/ml pepstatin A, 0.2 mM phenylmethylsulfonyl fluoride) and homogenized. The homogenate is centrifuged at 600 x g for 5 minutes. The particulate and cytosol fractions are separated by ultracentrifugation of the supernatant at 100,000 x g for 60 minutes. The nuclear fraction is obtained by resuspending the 600 x g pellet in sucrose solution (0.25 M sucrose/ 10 mM TRIS-HCl, pH 7.4/ 2 mM MgCl<sub>2</sub>) and recentrifuged at 600 x g. Equal amounts of protein from each fraction are applied to an SDS/10% polyacrylamide gel and blotted onto membranes. Western blot analysis is performed using CGDD anti-serum. The localization of CGDD is assessed by the intensity of the corresponding band in the nuclear fraction relative to the intensity in the other fractions. Alternatively, the presence of CGDD in cellular fractions is examined by fluorescence microscopy using a fluorescent antibody specific for CGDD.

Alternatively, CGDD activity may be demonstrated as the ability to interact with its associated Ras superfamily protein, in an *in vitro* binding assay. The candidate Ras superfamily

proteins are expressed as fusion proteins with glutathione S-transferase (GST), and purified by affinity chromatography on glutathione-Sepharose. The Ras superfamily proteins are loaded with GDP by incubating 20 mM Tris buffer, pH 8.0, containing 100 mM NaCl, 2 mM EDTA, 5 mM MgCl<sub>2</sub>, 0.2 mM DTT, 100  $\mu$ M AMP-PNP and 10  $\mu$ M GDP at 30°C for 20 minutes. CGDD is  
5 expressed as a FLAG fusion protein in a baculovirus system. Extracts of these baculovirus cells containing CGDD-FLAG fusion proteins are precleared with GST beads, then incubated with GST-Ras superfamily fusion proteins. The complexes formed are precipitated by glutathione-Sepharose and separated by SDS-polyacrylamide gel electrophoresis. The separated proteins are blotted onto nitrocellulose membranes and probed with commercially available anti-FLAG antibodies. CGDD  
10 activity is proportional to the amount of CGDD-FLAG fusion protein detected in the complex.

Alternatively, as demonstrated by Li and Cohen (Li, L. and S.N. Cohen (1995) Cell 85:319-329), the ability of CGDD to suppress tumorigenesis can be measured by designing an antisense sequence to the 5' end of the gene and transfecting NIH 3T3 cells with a vector transcribing this sequence. The suppression of the endogenous gene will allow transformed fibroblasts to produce  
15 clumps of cells capable of forming metastatic tumors when introduced into nude mice.

Alternatively, an assay for CGDD activity measures the effect of injected CGDD on the degradation of maternal transcripts. Procedures for oocyte collection from Swiss albino mice, injection, and culture are as described in Stutz et al., (*supra*). A decrease in the degradation of maternal RNAs as compared to control oocytes is indicative of CGDD activity. In the alternative,  
20 CGDD activity is measured as the ability of purified CGDD to bind to RNase as measured by the assays described in Example XVII.

Alternatively, an assay for CGDD activity measures syncytium formation in COS cells transfected with an CGDD expression plasmid, using the two-component fusion assay described in Mi (*supra*). This assay takes advantage of the fact that human interleukin 12 (IL-12) is a heterodimer  
25 comprising subunits with molecular weights of 35 kD (p35) and 40 kD (p40). COS cells transfected with expression plasmids carrying the gene for p35 are mixed with COS cells cotransfected with expression plasmids carrying the genes for p40 and CGDD. The level of IL-12 activity in the resulting conditioned medium corresponds to the activity of CGDD in this assay. Syncytium formation may also be measured by light microscopy (Mi et al., *supra*).

30 An alternative assay for CGDD activity measures cell proliferation as the amount of newly initiated DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding CGDD is transfected into quiescent 3T3 cultured cells using methods well known in the art. The transiently transfected cells are then incubated in the presence of [<sup>3</sup>H]thymidine or a radioactive DNA precursor such as [ $\alpha$ -<sup>32</sup>P]ATP. Where applicable, varying amounts of CGDD ligand are added to the  
35 transfected cells. Incorporation of [<sup>3</sup>H]thymidine into acid-precipitable DNA is measured over an

appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA and CGDD activity.

Alternatively, CGDD activity is measured by the cyclin-ubiquitin ligation assay (Townsend, F.M. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2362-2367). The reaction contains in a volume of 10  $\mu$ l, 40 mM Tris.HCl (pH 7.6), 5 mM Mg Cl<sub>2</sub>, 0.5 mM ATP, 10 mM phosphocreatine, 50  $\mu$ g of creatine phosphokinase/ml, 1 mg reduced carboxymethylated bovine serum albumin/ml, 50  $\mu$ M ubiquitin, 1  $\mu$ M ubiquitin aldehyde, 1-2 pmol <sup>125</sup>I-labeled cyclin B, 1 pmol E1, 1  $\mu$ M okadaic acid, 10  $\mu$ g of protein of M-phase fraction 1A (containing active E3-C and essentially free of E2-C), and varying amounts of CGDD. The reaction is incubated at 18 °C for 60 minutes. Samples are then separated by electrophoresis on an SDS polyacrylamide gel. The amount of <sup>125</sup>I- cyclin-ubiquitin formed is quantified by PHOSPHORIMAGER analysis. The amount of cyclin-ubiquitin formation is proportional to the activity of CGDD in the reaction.

Alternatively, an assay for CGDD activity uses radiolabeled nucleotides, such as [ $\alpha^{32}$ P]ATP, to measure either the incorporation of radiolabel into DNA during DNA synthesis, or fragmentation of DNA that accompanies apoptosis. Mammalian cells are transfected with plasmid containing cDNA encoding CGDD by methods well known in the art. Cells are then incubated with radiolabeled nucleotide for various lengths of time. Chromosomal DNA is collected, and radioactivity is detected using a scintillation counter. Incorporation of radiolabel into chromosomal DNA is proportional to the degree of stimulation of the cell cycle. To determine if CGDD promotes apoptosis, chromosomal DNA is collected as above, and analyzed using polyacrylamide gel electrophoresis, by methods well known in the art. Fragmentation of DNA is quantified by comparison to untransfected control cells, and is proportional to the apoptotic activity of CGDD.

Alternatively, cyclophilin activity of CGDD is measured using a chymotrypsin-coupled assay to measure the rate of cis to trans interconversion (Fischer, G. et al. (1984) *Biomed. Biochim. Acta* 43:1101-1111). The chymotrypsin is used to estimate the trans-substrate cleavage activity at Xaa-Pro peptide bonds, wherein the rate constant for the cis to trans isomerization can be obtained by measuring the rate constant of the substrate hydrolysis at the slow phase. Samples are incubated in the presence or absence of the immunosuppressant drugs CsA or FK506, reactions initiated by addition of chymotrypsin, and the fluorescent reaction measured. The enzymatic rate constant is calculated from the equation  $k_{app} = k_{H2O} + k_{enz}$ , wherein first order kinetics are displayed, and where one unit of PPIase activity is defined as  $k_{enz}$  (s<sup>-1</sup>).

Alternatively, cyclophilin activity of CGDD is monitored by a quantitative immunoassay that measures its affinity for stereospecific binding to the immunosuppressant drug cyclosporin (Quesniaux, V.F. et al. (1987) *Eur. J. Immunol.* 17:1359-1365). In this assay, the cyclophilin-cyclosporin complex is coated on a solid phase, with binding detected using anti-cyclophilin rabbit

antiserum enhanced by an antiglobulin-enzyme conjugate.

Alternatively, activity of CGDD is monitored by a binding assay developed to measure the non-covalent binding between FKBP<sub>s</sub> and immunosuppressant drugs in the gas phase using electrospray ionization mass spectrometry (Trepanier, D.J. et al. (1999) Ther. Drug Monit. 21:274-  
5 280). In electrospray ionization, ions are generated by creating a fine spray of highly charged droplets in the presence of a strong electric field; as the droplet decreases in size, the charge density on the surface increases. Ions are electrostatically directed into a mass analyzer, where ions of opposite charge are generated in spatially separate sources and then swept into capillary inlets where the flows are merged and where reactions occur. By comparing the charge states of bound versus unbound  
10 CGDD/immunosuppressive drug complexes, relative binding affinities can be established and correlated with *in vitro* binding and immunosuppressive activity.

Alternatively, a fluorescence monitoring assay for detecting activated Ras using RRP22 is as follows. The RRP22 binding domain (RRP22BD) of c-Raf1 (a kinase activated during reentry into meiosis) is synthesized from two unprotected peptide segments by native chemical ligation. Two  
15 fluorescent amino acids with structures based on the nitrobenz-2-oxa-1,3-diazole and coumaryl chromophores are incorporated close to the RRP22BD/RRP22-GTP binding surface followed by introduction of a C-terminal tag consisting of His(6). The  $K_D$  values for binding of the site-specifically modified proteins to Ras-GTP are compared to that of wild-type RBD. Ras-GTP is detected within the 100 nM range by immobilization of C-terminal His(6) tag-modified fluorescent  
20 RBD onto Ni-NTA-coated surfaces. Ras-GDP does not bind to the immobilized RBD, thus allowing discrimination between inactive and activated Ras (Becker, C. F. (2001) Chem. Biol. 8:243-252).

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of  
25 the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments.  
30 Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7722608	1	7722608CD1	34	7722608CB1
7505869	2	7505869CD1	35	7505869CB1
7502420	3	7502420CD1	36	7502420CB1
7506455	4	7506455CD1	37	7506455CB1
7506018	5	7506018CD1	38	7506018CB1
353017	6	353017CD1	39	353017CB1
5137816	7	5137816CD1	40	5137816CB1
7502151	8	7502151CD1	41	7502151CB1
7505983	9	7505983CD1	42	7505983CB1
7505986	10	7505986CD1	43	7505986CB1
3231075	11	3231075CD1	44	3231075CB1
7503516	12	7503516CD1	45	7503516CB1
7506179	13	7506179CD1	46	7506179CB1
1938744	14	1938744CD1	47	1938744CB1
5557436	15	5557436CD1	48	5557436CB1
7506178	16	7506178CD1	49	7506178CB1
7506235	17	7506235CD1	50	7506235CB1
1302184	18	1302184CD1	51	1302184CB1
7506232	19	7506232CD1	52	7506232CB1
2585358	20	2585358CD1	53	2585358CB1
3961495	21	3961495CD1	54	3961495CB1
7500801	22	7500801CD1	55	7500801CB1
7506414	23	7506414CD1	56	7506414CB1
7506415	24	7506415CD1	57	7506415CB1
72192179	25	72192179CD1	58	72192179CB1
7505908	26	7505908CD1	59	7505908CB1
6590147	27	6590147CD1	60	6590147CB1
6828539	28	6828539CD1	61	6828539CB1
7170321	29	7170321CD1	62	7170321CB1

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7505918	30	7505918CD1	63	7505918CB1
7505935	31	7505935CD1	64	7505935CB1
4225965	32	4225965CD1	65	4225965CB1
7495594	33	7495594CD1	66	7495594CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7722608CD1	g6708478	0.0	[Mus musculus] formin-like protein. Yayoshi-Yamamoto, S. et al. (2000) FRL, a novel formin-related protein, binds to rac and regulates cell motility and survival of macrophages. Mol. Cell. Biol. 20:6872-6881.
		614055[Frls	0.0	[Mus musculus][Cytoplasmic] Protein with weak similarity to diaphanous and diaphanous-related proteins, which are profilin ligands and regulators of polymerization of actin and which are associated with nonsyndromic deafness (DFNA1) in humans. Yayoshi-Yamamoto, S. et al. (2000) FRL, a novel formin-related protein, binds to Rac and regulates cell motility and survival of macrophages. Mol. Cell. Biol. 20: 6872-6881.
2	7505869CD1	g18568346	0.0	[Homo sapiens] atypical PKC isotype-specific interacting protein long variant. Gao, L. et al. (2002) Multiple splice variants of Par3 and of a novel related gene, Par3L, produce proteins with different binding properties. Gene 294:99
		332858[Rn.31803	0.0	[Rattus norvegicus][Endoplasmic reticulum; Cytoplasmic; Plasma membrane; Cell junction] Atypical PKC isotype-specific interacting protein, localizes to tight junctions, binds to atypical protein kinase C isozymes and forms a complex that may be involved in the polarity of epithelial cells. Lin, D. et al. (2000) A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and aPKC signalling and cell polarity. Nat. Cell Biol. 2:540-547.
		606222[PARD3	0.0	[Homo sapiens][Plasma membrane; Cell junction] Multi-PDZ protein that is essential for asymmetric cell division and polarized growth, may have a role in the formation of tight junctions at epithelial cell-cell contacts. Joberty, G. et al. (2000) The cell-polarity protein Par6 links Par3 and atypical protein kinase C to Cdc42. Nat. Cell Biol. 2: 531-539.



Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
3	7502420CD1	g13182960	3.2E-66	[Mus musculus] breast metastasis suppressor 1-like protein.
4	7506455CD1	g13097315	1.2E-101	[Homo sapiens] ovarian carcinoma immunoreactive antigen.
5		g13539559	3.3E-12	[Homo sapiens] SH3BCRL3-like protein
6	353017CD1	g2970644	2.1E-67	[Gallus gallus] Xin Wang, D.Z. et al. (1996) Differential displaying of mRNAs from the atrioventricular region of developing chicken hearts at stages 15 and 21. Frontiers in bioscience [computer file] : a journal and virtual library. 1:A1-A15 Wang, D.Z. et al. (1999) Requirement of a novel gene, Xin, in cardiac morphogenesis. Development 126:1281-1294
		324048[Xin	6.6E-35	[Mus musculus] Protein with weak similarity to yeast Muc1p cell surface flocculin, required for invasive and pseudohyphal growth Wang, D. Z. et al. (1999) Requirement of a novel gene, Xin, in cardiac morphogenesis. Dev Suppl 126:1281-94
7	5137816CD1	g9887326	0.0	[Homo sapiens] angiominin Troyanovsky, B. et al. (2001) Angiominin. An angiostatin binding protein that regulates endothelial cell migration and tube formation. J. Cell Biol. 152:1247-1254
		425080[KIAA1071	2.7E-243	[Homo sapiens] Protein of unknown function, has weak similarity to a region of murine NfH heavy subunit of neurofilament, which is an intermediate filament that contains 48 repeats of a Lys-Ser-Pro motif that is a kinase recognition site Kikuno, R. et al. (1999) Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res. 6:197-205
		607284[Lccp	5.10E-107	[Mus musculus] Protein of unknown function, has strong similarity to uncharacterized human LCCP

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
8	7502151CD1	g9837581	2.8E-26	[Drosophila melanogaster] WIBG Ohlstein,B. et al. (2000) The drosophila cystoblast differentiation factor, benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with bag-of-marbles. Genetics 155:1809-1819
		637330 orf6.3786	1.7E-11	[Candida albicans] Protein of unknown function, has moderate similarity to S. cerevisiae Ygr054p, which may play a role in signal transduction Huang, S. et al. (1987) Specificity of cotranslational amino-terminal processing of proteins in yeast. Biochemistry 26:8242-8246
9	7505983CD1	g189422	0.0	[Homo sapiens] proliferating cell nuclear protein P120 Fonagy,A. et al. (1989) Cloning of the cDNA and sequence of the human proliferating cell nuclear protein P120. Cancer Commun. 1:243-251
		343564 NOL1	0.0	[Homo sapiens][RNA-binding protein][Nuclear nucleolus; Nuclear] Nucleolar protein 1, a proliferation-associated antigen expressed during early G1 and S phases of the cell cycle and following DNA damage; transforms NIH3T3 cells when overexpressed; binds 60-80S pre-RNP particles and nuclear proteins MCRS1 and NPM1 Saijo, Y. et al. (1993) The effect of antisense p120 construct on p120 expression and cell proliferation in human breast cancer MCF-7 cells. Cancer Lett. 68:95-104
		8252 NOP2	1.7E-119	[Saccharomyces cerevisiae][RNA-binding protein][Nuclear nucleolus; Nuclear] Nucleolar protein required for pre-rRNA processing and 60S ribosome synthesis, has strong similarity to human proliferation-associated p120 nucleolar antigen Van Dyck, L., et al. (1994) Analysis of a 17.4 kb DNA segment of yeast chromosome II encompassing the ribosomal protein L19 as well as proteins with homologies to components of the hnRNP and snRNP complexes and to the human proliferation-associated p120 antigen. Yeast 10:1663-1673
10	7505986CD1	g189422	3.3E-51	[Homo sapiens] proliferating cell nuclear protein P120 Fonagy,A. et al. (1989) Cloning of the cDNA and sequence of the human proliferating cell nuclear protein P120. Cancer Commun. 1:243-251

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
10		343564 NOL1	2.6E-52	[Homo sapiens][RNA-binding protein][Nuclear nucleolus; Nuclear] Nucleolar protein 1, a proliferation-associated antigen expressed during early G1 and S phases of the cell cycle and following DNA damage; transforms NIH3T3 cells when overexpressed; binds 60-80S pre-RNP particles and nuclear proteins MCRS1 and NPM1 Fonagy, A. et al. (1994) Effect of nucleolar P120 expression level on the proliferation capacity of breast cancer cells. Cancer Res. 54:1859-1864
11	3231075CD1	g3095098	0.0	[Homo sapiens] truncated EVI5 Roberts, T. et al. (1998) NB4S, a member of the TBC1 domain family of genes, is truncated as a result of a constitutional t(1;10)(p22;q21) chromosome translocation in a patient with stage 4S neuroblastoma. Hum. Mol. Genet. 7:1169-1178
		342650 EVI5	0.0	[Homo sapiens][Small molecule-binding protein] Ecotropic viral integration site 5, may regulate cell growth and differentiation; translocation of the corresponding gene is associated with stage 4S neuroblastoma Roberts, T. et al. (1998) NB4S, a member of the TBC1 domain family of genes, is truncated as a result of a constitutional t(1;10)(p22;q21) chromosome translocation in a patient with stage 4S neuroblastoma. Hum. Mol. Genet. 7:1169-1178
		327342 Evi5	3.1E-290	[Mus musculus] Protein with similarity to human NB4S, which may regulate cell growth and differentiation; corresponding gene is a common site of retroviral integration in AKXD T-cell lymphomas, and translocation of human EVI5 is associated with stage 4S neuroblastoma Liao, X. et al. (1995) Evi-5, a common site of retroviral integration in AKXD T-cell lymphomas, maps near Gfi-1 on mouse chromosome 5. J. Virol. 69:7132-137
12	7503516CD1	g9309467	2.5E-22	[Mus musculus] leucine-rich glioma-inactivated 1 protein precursor Somerville, R.P.T. et al. (2000) Identification of the promoter, genomic structure, and mouse ortholog of LGH1. Mamm. Genome 11:622-627

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
12		341690 LGI1	2.2E-23	[Homo sapiens] Leucine-rich glioma inactivated 1, a member of the LRR-protein superfamily that contains four leucine-rich repeats, may function in neurogenesis, may be a tumor suppressor; gene rearrangement and low expression are associated with malignant glioma Chernova, O. B. et al. (1998) A novel gene, LGI1, from 10q24 is rearranged and downregulated in malignant brain tumors. Oncogene 17:2873-2881
		607582 Lgi1	2.2E-23	[Mus musculus] Leucine-rich glioma inactivated 1, contains four leucine-rich repeats, putative glial tumor suppressor; gene rearrangement and low expression of the human LGI1 gene are associated with malignant gliomas Somerville, R.P.T. et al. (2000) Identification of the promoter, genomic structure, and mouse ortholog of LGI1. Mamm. Genome 11:622-627
13	7506179CD1	g1297330	2.5E-297	[Homo sapiens] DOC-2 Mok, S.C. et al. (1994) Molecular cloning of differentially expressed genes in human epithelial ovarian cancer. Gynecol. Oncol. 52:247-252 Mok, S.C. et al. (1998) DOC-2, a candidate tumor suppressor gene in human epithelial ovarian cancer. Oncogene 16:2381-2387
		334984 DAB2	2.2E-298	[Homo sapiens] Disabled homolog 2, binds MADH2 and MADH3, competes with SOS1 for binding to GRB2 and uncouples MAPK activation from FOS expression, negatively regulates cell proliferation, involved in cell adhesion; expression is lost in ovarian tumors Fazili, Z., et al. (1999) Disabled-2 inactivation is an early step in ovarian tumorigenicity. Oncogene 18:3104-3113

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
13		664905 Dab2	1.2E-237	[Rattus norvegicus] Protein that slows cell growth when overexpressed, expression is likely regulated by androgens, has strong similarity to murine Mm.4544, which is phosphorylated when cells are stimulated by a mitogenic factor Tseng, C. P., et al. (1998) Regulation of rat DOC-2 gene during castration-induced rat ventral prostate degeneration and its growth inhibitory function in human prostatic carcinoma cells. Endocrinology 139:3542-3553
16	7506178CD1	g3551742	6.3E-160	[Homo sapiens] HT-1080 protein Onisto, M. et al. (1998) Expression study on D123 gene product: evidence for high positivity in testis. Exp. Cell Res. 242:451-459
		342970 D123	6.2E-160	[Homo sapiens] D123 gene product, cytoplasmic protein with a putative role in cell cycle progression, highly expressed in the testes Okuda, A. and Kimura, G. (1996) An amino acid change in novel protein D123 is responsible for temperature-sensitive G1-phase arrest in a mutant of rat fibroblast line 3Y1. Exp Cell Res 223:242-249
		331628 C10	1.4E-95	[Rattus norvegicus] D123 gene product, protein with a putative role in cell cycle progression Okuda, A. and Kimura, G. (1996) An amino acid change in novel protein D123 is responsible for temperature-sensitive G1-phase arrest in a mutant of rat fibroblast line 3Y1. Exp Cell Res 223:242-249
17	7506235CD1	g406263	6.3E-98	[Mus musculus] T10 Halford, S. et al. (1993) Isolation of a gene expressed during early embryogenesis from the region of 22q11 commonly deleted in DiGeorge syndrome. Hum. Mol. Genet. 2:1577-1582
18	1302184CD1	g5478318	0.0	[Homo sapiens] SET-binding protein (SEB) Minakuchi, M. et al. (2001) Identification and characterization of SEB, a novel protein that binds to the acute undifferentiated leukemia-associated protein SET. Eur. J. Biochem. 268:1340-1351

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
18		742852 SETBP1	0.0	[Homo sapiens] SET binding protein 1, a nuclear protein that binds SET, which is encoded by a gene at a translocation breakpoint associated with acute undifferentiated leukemia Minakuchi, M. et al. (2001) Identification and characterization of SEB, a novel protein that binds to the acute undifferentiated leukemia-associated protein SET. Eur J Biochem 268:1340-1351
19	7506232CD1	g1469920	4.6E-92	[Homo sapiens] D53 Byrne, J.A. et al. (1995) A screening method to identify genes commonly overexpressed in carcinomas and the identification of a novel complementary DNA sequence. Cancer Res. 55:2896-2903 Byrne, J.A. et al. (1996) Definition of the tumor protein D52 (TPD52) gene family through cloning of D52 homologues in human (hD53) and mouse (mD52). Genomics 35:523-532
		338644 TPD52L 1	4.0E-93	[Homo sapiens] Tumor protein D52-like 1, a member of the TPD52 gene family consisting of a coiled-coil domain and PEST domains, may form homodimers or heterodimers with TPD52, and may play a role in cell proliferation Nourse, C. R. et al. (1998) Cloning of a third member of the D52 gene family indicates alternative coding sequence usage in D52-like transcripts. Biochim. Biophys. Acta 1443:155-168
		323536 Tpd52l1	1.9E-84	[Mus musculus] Tumor protein D52-like 1, contains a coiled-coil domain, interacts in homo- and heteromeric fashions with other D52-like proteins, and may have a role in controlling cell proliferation Byrne, J. A. et al. (1998) Identification of homo- and heteromeric interactions between members of the breast carcinoma-associated D52 protein family using the yeast two- hybrid system. Oncogene 16:873-881
20	2585358CD1	g3255991	1.7E-16	[Homo sapiens] NY-ESO-1 protein. Lethe, B. et al. (1998) Int. J. Cancer 76:903-908.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
20		334864 CTAG1	1.4E-17	[Homo sapiens] Cancer testis antigen 1, a tumor antigen aberrantly expressed in a number of cancers, may be a candidate vaccine target to elicit a tumor specific cellular and humoral defense response against cancer.
21	3961495CD1	g9295520	9.0E-59	Lethe, B. et al. (1998) Int. J. Cancer 76:903-908. [Rattus norvegicus] androgen receptor-related apoptosis-associated protein CBL27.
		735175  KIAA1093	2.1E-149	[Homo sapiens][RNA-binding protein] Protein containing an RNA recognition motif (RRM, RBD, or RNP).
		618262 DSPP	3.1E-13	[Homo sapiens][Structural protein][Extracellular matrix (cuticle and basement membrane)]; Extracellular (excluding cell wall) Dentin sialophosphoprotein (phosphoryn), an extracellular matrix protein of dentin in teeth, contains RGD sequence motifs, may function in dentin mineralization; gene mutation is associated with dentinogenesis imperfecta Shields type II.
22	7500801CD1	g21901935	0.0	Rowe, P. S. et al. (2000) Genomics 67:54-68. [Mus musculus] leucine-rich glioma-inactivated 1 protein precursor.
		341690 LGI1	1.7E-117	Scheel, H. et al. (2002) A common protein interaction domain links two recently identified epilepsy genes. Hum. Mol. Genet. 11:1757-1762. [Homo sapiens] Leucine-rich glioma inactivated 1, a member of the LRR-protein superfamily that contains four leucine-rich repeats, may function in neurogenesis, may be a tumor suppressor; gene rearrangement and low expression are associated with malignant glioma.
		662837 Slit1	8.1E-17	Chernova, O. B. et al. (1998) Oncogene 17:2873-2881. [Rattus norvegicus][Extracellular (excluding cell wall)] Slit, a secreted protein that may act to guide the direction of neuronal migration in the developing olfactory system, expressed only in fetal and adult forebrain neurons.
23	7506414CD1	g18032008	0.0	Rochet, N. et al. (1999) Int. J. Cancer 82:282-285. [Drosophila melanogaster] Scribble.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
23		599864 ERBB2IP	2.3E-118	[Homo sapiens][Basolateral plasma membrane; Extracellular matrix (cuticle and basement membrane); Basement membrane (extracellular matrix)] ERBB2-interacting protein ERBIN, contains 16 leucine-rich repeats and a PDZ domain, binds ERBB2, BPAG1 hemidesmosomal protein, and integrin beta B4 subunit (ITGB4), functions as an adaptor protein that may link ERBB2 signaling to hemidesmosome assembly. Jaulin-Bastard, F. et al. (2001) J. Biol. Chem. 276:15256-63.
		756686 LOC117284	3.4E-111	[Rattus norvegicus] Densin-180, a postsynaptic density protein that binds calcium/calmodulin-dependent protein kinase II. Strack, S. et al. (2000) J. Biol. Chem. 275:25061-25064.
		692090 LANO	3.7E-73	[Homo sapiens] LAP and no PDZ domain, a member of the LAP family that lacks a PDZ domain, a basolateral epithelial cell protein which binds to the PDZ domain of MAGUK proteins and indirectly binds Erbin (ERBB2IP), may participate in epithelial tissue homeostasis. Saito, H. et al. (2001) J. Biol. Chem. 276:32051-32055.
24	7506415CD1	g18032008	0.0	[Drosophila melanogaster] Scribble.
		750792 KIAA0147	0.0	[Homo sapiens] Protein containing PDZ (DHR, GLGF) domains, which target signaling proteins to membranes, contains leucine rich repeats, which mediate protein-protein interactions.
		599864 ERBB2IP	6.0E-118	[Homo sapiens][Basolateral plasma membrane; Extracellular matrix (cuticle and basement membrane); Basement membrane (extracellular matrix)] ERBB2-interacting protein ERBIN, contains 16 leucine-rich repeats and a PDZ domain, binds ERBB2, BPAG1 hemidesmosomal protein, and integrin beta B4 subunit (ITGB4), functions as an adaptor protein that may link ERBB2 signaling to hemidesmosome assembly. Borg, J. P. et al. (2000) Nat. Cell. Biol. 2:407-414.



Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
25	72192179CD1	g2970644	3.6E-13	[Gallus gallus] Xin. Wang, D.Z. et al. (1999) Development (Cambridge, England) 126:1281-1294.
26	7505908CD1	g2330011	4.6E-54	[Mus musculus] mm-Mago. Newmark, P.A. et al. (1997) Development 124:3197-3207.
		336336[MAGOH	3.9E-55	[Homo sapiens] Mago-nashi homolog, may play a role in RNA localization and germ cell development. Zhao, X. F. et al. (1998) Genomics 47:319-322.
		585247[Mago	3.9E-55	[Mus musculus] Mago-nashi homolog, serum-inducible protein that may play a role in germ cell development. Zhao, X. F. et al. (1998) Genomics <i>supra</i> .
27	6590147CD1	g13436392	3.0E-177	[Homo sapiens] cyclin I
		428176[CCNI	2.5E-178	[Homo sapiens][Regulatory subunit; Cyclin] Cyclin I, member of the cyclin family of CDK kinase regulatory subunits Nakamura, T. et al. (1995) Cyclin I: a new cyclin encoded by a gene isolated from human brain. Exp Cell. Res. 221:534-542
		587655[Ccni	1.4E-166	[Mus musculus][Regulatory subunit; Cyclin] Protein with strong similarity to human CCNI (Cyclin I), which is a member of the cyclin family of CDK kinase regulatory subunits Jensen, M. R. et al. (2000) In vivo expression and genomic organization of the mouse cyclin I gene (Ccni). Gene 256:59-67
28	6828539CD1	g9944942	6.1E-131	[Homo sapiens] DRC3 Wu, K. et al. (1999) Cloning and expression analyses of down-regulated cDNA C6-2A in human esophageal cancer. Chung-Hua I Hsueh I Chuan Hsueh Tsa Chih 16: 325-327

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
28		599224  FLJ20258	2.1E-194	[Homo sapiens] Protein containing an Src homology 3 (SH3) domain, has low similarity to EPS8 (epidermal growth factor receptor pathway substrate), which is tyrosine phosphorylated by epidermal growth factor receptor (EGFR) and enhances EGF-dependent mitogenic signals
		340492 EPS8	5.3E-94	[Homo sapiens][Receptor (signalling)][Nuclear] Epidermal growth factor receptor pathway substrate 8, SH3 containing protein that is tyrosine phosphorylated by epidermal growth factor receptor (EGFR) and enhances EGF-dependent mitogenic signals, has a role in normal and neoplastic cell proliferation
		586475 Eps8	9.6E-94	Matoskova, B. et al. (1995) Constitutive phosphorylation of eps8 in tumor cell lines: relevance to malignant transformation. Mol. Cell. Biol. 15:3805-3812
				[Mus musculus][Receptor (signalling)][Nuclear] Epidermal growth factor receptor pathway substrate 8, an adaptor that enhances EGF-induced mitogenesis and mediates PDGF (Pdgfb) induced Rac1 activation, actin reorganization and membrane ruffling; human EPS8 has a role in neoplastic cell proliferation
				Maa, M. C. et al. (2001) Overexpression of p97Eps8 leads to cellular transformation: implication of pleckstrin homology domain in p97Eps8-mediated ERK activation. Oncogene 20:106-112
29	717032 CD1	g14250656 328058 Rn.2140	3.5E-135 2.0E-187	[Homo sapiens] CDC37 (cell division cycle 37, S. cerevisiae, homolog)
				[Rattus norvegicus] Cell division cycle 37 homolog, subunit of a complex that includes the molecular chaperone Hsp90 and Cdk4 (Rn.6115), may play a role in the cell division cycle
				Ozaki, T. et al. (1995) Molecular cloning and cell cycle-dependent expression of a novel gene that is homologous to cdc37. DNA Cell Biol. 14:1017-1023

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
29		428266 CDC37	3.0E-136	[Homo sapiens][Chaperones; Small molecule-binding protein] Cell division cycle 37 homolog, a cell cycle regulator that associates with molecular chaperone Hsp90, which targets it to protein kinase CDK4 to promote CDK4 stability, plays a role in heme biosynthesis Stepanova, L. et al. (2000) Induction of human Cdc37 in prostate cancer correlates with the ability of targeted Cdc37 expression to promote prostatic hyperplasia. Oncogene 19:2186-2193
30	7505918CD1	g1679792	0.0	[Danio rerio] Allele: hi4 Allende, M.L. et al. (1996) Insertional mutagenesis in zebrafish identifies two novel genes, pescadillo and dead eye, essential for embryonic development. Genes Dev. 10: 3141-3155
		569934  KIAA0095	0.0	[Homo sapiens][Nuclear pore; Nuclear] Nuclear pore protein, physically interacts with 205-kDa protein and the nucleoporin p62 Grandi, P. et al. (1997) Nup93, a vertebrate homologue of yeast Nic96p, forms a complex with a novel 205-kDa protein and is required for correct nuclear pore assembly. Mol. Biol. Cell 8:2017-2038
		9717 NIC96	4.1E-53	[Saccharomyces cerevisiae][Nuclear import/export protein][Nuclear; Nuclear pore] Nuclear pore protein (nucleoporin), acts in a complex with Nsp1p, Nup57p, and Nup49p Zabel, U. et al. (1996) Nic96p is required for nuclear pore formation and functionally interacts with a novel nucleoporin, Nup188p. J. Cell. Biol. 133:1141-1152
31	7505935CD1	g13399213	2.3E-36	[Mus musculus] putative cxorf-5 related protein Ferrante, M.I. et al. (2001) Identification of the gene for oral-facial-digital type I syndrome. Am. J. Hum. Genet. 68:569-576

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
31		334922 OFD1	0.0	[Homo sapiens] Protein with many coiled-coiled domains that is expressed ubiquitously and escapes X inactivation de Conciliis, L. et al. (1998) Characterization of Cxorfs (71-7A), a novel human cDNA mapping to Xp22 and encoding a protein containing coiled-coil alpha-helical domains. Genomics 51:243-250
		590383 Myh9	6.9E-20	[Rattus norvegicus][Motor protein; Hydrolase; ATPase] Non-muscle myosin heavy chain 9, a motor protein that is predicted to provide force for cytokinesis and cell adhesion; mutation of human MYH9 gene is associated with May Hegglin anomaly, Fechtner, or Sebastian syndrome and non-syndromic deafness Lalwani, A. K. et al.(2000) Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in nonmuscle myosin MYH9. Am. J. Hum. Genet. 67:1121-1128
32	4225965CD1	g14583268	3.6E-188	[Homo sapiens] cytoplasmic protein, Ropp120 (restrictedly overexpressed proliferation-associated protein) Heidebrecht, H.J. et al. (2000) Cloning and localization of C2orf2(ropp120), a previously unknown WD repeat protein. Genomics 68:348-350
		428754 EMAP-2	4.8E-175	[Homo sapiens][Cytoplasmic; Cytoskeletal] Echinoderm microtubule-associated protein 2, a member of a family of proteins with similarity to an echinoderm microtubule-associated protein, may be involved in vision and hearing Lepley, D. M., et al. (1999) Sequence and expression patterns of a human EMAP-related protein-2 (HuEMAP-2). Gene 237:343-349
		244700 F38A6.2	1.6E-66	[Caenorhabditis elegans] Protein with strong similarity to human echinoderm microtubule-associated-like protein EMAPL (Hs.12451) Suprenant, K. A. et al. (2000) Conservation of the WD-repeat, microtubule-binding protein, EMAP, in sea urchins, humans, and the nematode C. elegans. Dev. Genes Evol. 210:2-10
33	7495594CD1	g12060830	2.6E-263	[Homo sapiens] serologically defined breast cancer antigen NY-BR-38

Table 2

Polypeptide SEQ ID NO:	Incye Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
33		662671 Polydom	0.0	[Mus musculus] Polydomain protein, a secreted protein with many domains, including pentraxin, von Willebrand factor type A, EGF, and complement control protein module domains, may play a role in cell adhesion and immunity Gilges, D. et al. (2000) Polydom: a secreted protein with pentraxin, complement control protein, epidermal growth factor and von willebrand factor A domains. Biochem. J. 352 Pt 1:49-59
		690824  POLYDOM	1.2E-235	[Homo sapiens] Polydomain protein, a likely ortholog of mouse Polydom which is a secreted protein with many domains, including pentraxin, von Willebrand factor type A, EGF, and complement control protein module domains Gilges, D. et al. (2000) Polydom: a secreted protein with pentraxin, complement control protein, epidermal growth factor and von willebrand factor A domains. Biochem. J. 352 Pt 1:49-59

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7722608CD1	1104	S57 S89 S97 S127 S173 S175 S184 S211 S222 S318 S367 S456 S474 S512 S553 S624 S675 S779 S827 S831 S850 S851 S969 S1059 S1086	N420 N849	Formin Homology 2 Domain: A632-D1067	HMMER_PFAM
			T203 T365 T412 T442 T708 T721 T758 T766 T873 T878 T929 T962 T1082 Y1057		PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPIANCHOR BRAIN MAJOR PD001091: S512-P643	BLAST_PRODROM
					Related Lymphocyte Formin 117.5 KD PROTEIN PD127019: L626-S697	BLAST_PRODROM
					PROTEIN DEVELOPMENTAL FORMIN LIMB DEFORMITY NUCLEAR ALTERNATIVE SPLICING CELL DIAPHANOUS PD003542: L814-P1054	BLAST_PRODROM
					FORMIN; DM04565  Q05860 176-1467: V384-E1003  Q05859 5-1205: V384-F986  Q05858 1-1212: E399-E932	BLAST_DOMO
					FORMIN: DM05091  P41832 835-1744: A533-R931	BLAST_DOMO

Table 3

SEQ ID NO.	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	7505869CD1	1299	S103 S124 S134 S141 S146 S167 S228 S247 S272 S350 S352 S359 S408 S503 S513 S584 S665 S695 S700 S722 S760 S761 S807 S809 S820 S868 S904 S935 S938 S1045 S1059 S1082 S1086 S1092 S1135 S1171 S1188 S1195 S1251 S1255	N92 N139 N188 N245 N371 N574 N588 N902 N1087 N1174	PDZ domain (Also known as DHR or GLGF): N441-Q527, E570-R663, K253-A340	HMMER_PFAM
					PDZ domain proteins Als PF00595: L485-N495	BLIMPS_PFAM
			T223 T238 T433 T448 T456 T514 T549 T634 T641 T685 T827 T845 T927 T945 T968 T975 T1119 Y179 Y368 Y725 Y913 Y1023 Y1264		PROTEIN SH3 DOMAIN REPEAT PD00289: G488-G501	BLIMPS_PRODOM
					PROTEIN DOMAIN PROTEASE PHOSPHATASE SH3 REPEAT PDZ TYROSINE PRECURSOR HYDROLASE PD000073: E570-R662	BLAST_PRODOM
					ATP/GTP-binding site motif A (P-loop): G496-S503, G627-T634	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	7502420CD1	433	S53 S102 S115 S138 S139 S140 S149 S157 S176 S195 S240 S258 S269 S298 S363 S429 T122 T218 T336 Y372 Y378		Cell attachment sequence: R116-D118	MOTIFS
4	7506455CD1	199	S128 S137 S172 T33 T102 T196	N174	Cytosolic domains: M1-E46, K101-E199 Transmembrane domains: S47-S68, Y78-V100 Non-cytosolic domain: K69-K77	TMHMMER
5	7506018CD1	37	S14 T29			
6	353017CD1	312	S82 S119 S142 S206 S249 T63 T95 T101 T179 T284		XIN PROTEIN PD044293: D15-K308, E6-I294, R58-C311	BLAST_PRODOM
					XIN HYPOTHETICAL PROTEIN PD044454: Q16-D261, D98-D302	BLAST_PRODOM
					Cell attachment sequence: R58-D60	MOTIFS
7	5137816CD1	993	S6 S82 S84 S122 S138 S210 S237 S392 S393 S456 S472 S546 S678 S683 S728 S781 T17 T442 T469 T473 T538 T588 T626 T630 T773 T775 Y365 Y436	N560 N625 N782	PROTEIN CALPHOTIN CALCIUM BINDING CYT ADHERENCE HIGH MOLECULAR WEIGHT ACCESSORY STRUCTURAL FILAMENTOUS PD016116: P787-P964  MOUSE NEUROFILAMENT TRIPLET H PROTEIN; DM04498P19246 429-715: K723-P964, K452-K494	BLAST_PRODOM  BLAST_DOMO



Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	7502151CD1	204	S55 S74 S117 T12 T19 T25 T86 T148			
9	7505983CD1	630	S40 S44 S111 S258 S315 S380 S383 S406 S464 S493 S520 S525 S526 S608 S619 T64 T129 T134 T137 T295 T323 T338 T360 T384 T423 T513 T557 T594	N321 N436	NOL1/NOP2/sun family: F118-K403	HMMER_PFAM
					NOL1/NOP2/sun family proteins BL01153: K325-L350, I206-M220, G228-G251, F272-G285	BLIMPS_BLOCKS
					PROTEIN FMU LONG NUCLEOLAR SUN P120 NUCLEAR PROLIFERATING CELL ANTIGEN 450AA PD003196: A121-K403	BLAST_PRODUM
					PROLIFERATING CELL NUCLEOLAR ANTIGEN P120 PROLIFERATION ASSOCIATED PROTEIN NUCLEAR PD119014: M1-E117	BLAST_PRODUM
					PROLIFERATING CELL NUCLEOLAR ANTIGEN P120 PROLIFERATION ASSOCIATED PROTEIN NUCLEAR PD124159:K442-I507	BLAST_PRODUM
					NOL1/NOP2/FMU FAMILY DM02871  P46087 189-575: E117-M394  P40991 155-537: S111-M394, E113-K136  P36929 5-428: K105-G361	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9					NOL1/NOP2/FMU FAMILY DM07271 P46087 577-645: D395-S464	BLAST_DOMO
10	7505986CD1	129	S40 S44 T64		NOL1/NOP2/sun family signature: F272-G283 PROLIFERATING CELL NUCLEOLAR ANTIGEN P120 PROLIFERATION ASSOCIATED PROTEIN NUCLEAR PD119014: MI-G111	MOTIFS BLAST_PRODROM
11	3231075CD1	607	S14 S102 S109 S113 S128 S196 S235 S283 S405 S473 S497 T3 T20 T209 T417 T547 T568 T576 Y415	N107 N126 N495	TBC domain: V160-L370	HMNER_PFAM
					Probable rabGAP domain p PF00566: I202-P211, Y243-A248, G252-M256	BLIMPS_PFAM
					EV15 HOMOLOG TRUNCATED EV15 ECOTROPIC VIRAL INTEGRATION SITE PD022955: M1-V156	BLAST_PRODROM
					EV15 HOMOLOG TRUNCATED EV15 ECOTROPIC VIRAL INTEGRATION SITE COSMID F01G12 PD075221: R431-P554	BLAST_PRODROM
					PROTEIN CHROMOSOME TRANSMEMBRANE CELL DIVISION I OF ONCOGENE COSMID SIMILAR PD001799: E225-M375, K185-L236, P75-1136	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11					YMR192W; SPAC4G8.04; YOL112W; MEMBRANE; DM01737  S61015 373-635: E97-R354  Q10496 503-811: S47-F353  S50920 216-482: V118-R354  S62481 395-698: K150-R354 Binding-protein-dependent transport systems inner membrane comp. sign: M231-P259 signal_cleavage: M1-A19	BLAST_DOMO
12	7503516CD1	105	S89		Signal Peptide: M1-A19, M1-V16, M1-R21, M1-P22, M1-G25	SPSCAN HMIMER
13	7506179CD1	552	S24 S208 S231 T203 T221 T420 Y50	N96 N191	Phosphotyrosine interaction domain (PTB/PID): Y50-E176	HMIMER_PFAM
					DOC2 PHOSPHORYLATION PROTEIN MITOGEN RESPONSIVE PHOSPHOPROTEIN ALTERNATIVE SPLICING ISOFORM DIFFERENTIALLY EXPRESSED PD025095: S220-A552, A185-L268	BLAST_PRODUM
					PROTEIN ALTERNATIVE PHOSPHORYLATION SPLICING DOC2 ISOFORM INITIATION REGULATION NEURONAL MUNC181 PD006376: P61-K175	BLAST_PRODUM
					DOC2 PROTEIN PHOSPHORYLATION MITOGEN RESPONSIVE PHOSPHOPROTEIN ALTERNATIVE SPLICING ISOFORM DIFFERENTIALLY EXPRESSED PD024160: Q13-D59	BLAST_PRODUM
					Cell attachment sequence: R64-D66	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13					pfkB family of carbohydrate kinases signature 1: A293-G316	MOTIFS
14	1938744CD1	54	S18			
15	5557436CD1	479	S170 S403 S445 T11 T47 T80 T135 T146 T224 T386		UBX domain: T358-A438	HMMER_PFAM
16	7506178CD1	295	S47 S56 S60 S70 S121 S234 S262 S273 T22 T79 T126 T219 T249 T274	N106 N119	signal_cleavage: M1-S56	SPSCAN
					PROTEIN D123 HT1080 CELL CYCLE CHROMOSOME XII COSMID PD024673: M1-1188, V190-D295	BLAST_PRODOM
17	7506235CD1	214	S48 S65 S130 S194 S201 T189 T191		PROTEIN REGION R186.1 SER/THRRICH T10 DCCR ASN2PHB1 INTERGENIC PD024111: M1-G89, D82-L212	BLAST_PRODOM
18	1302184CD1	1596	S9 S24 S65 S73 S77 S92 S102 S159 S241 S269 S304 S333 S337 S338 S376 S391 S423 S453 S515	N97 N386 N395 N714 N798 N875 N876 N1026 N1169 N1416	ATHILA ORF-1 family: P637-A689	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18			S551 S716 S754 S847 S859 S877 S888 S893 S910 S932 S944 S970 S1011 S1015 S1144 S1187		AT hook motif: K1016-T1028, K584-L596, K1451-Q1463	HMMER_PFAM
			S1191 S1201 S1235 S1264 S1266 S1272 S1298 S1302 S1325 S1349 S1394 S1422 S1483 S1590		ACROSIN DM03631  P12978 10-163: P1489-E1578  P48038 288-430: A1516-R1549	BLAST_DOMO
			T109 T110 T132 T151 T165 T205 T224 T248 T508 T533 T572 T600 T620 T662		FIBRILLAR COLLAGEN CARBOXYL-TERMINAL DM00042 A41132 43-133: P1520-P1572	BLAST_DOMO
			T671 T720 T818 T826 T1260 T1273 T1308 T1360 T1423 T1462 T1547 Y994		Sugar transport proteins signature 2: L1170-K1195	MOTIFS
19	7506232CD1	191	S32 S58 S86 S122 S136 S147 S161 S181 T17 T52 T133 T184	N82	PROTEIN D52 TUMOR D53 N8 CSPP28 MD52 R10 HD53 HD54+INS2 PD008422: M1-R182, E2-L189	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19					Tumor protein D52 DM06235[P55327][1-183: M1-E186	BLAST_DOMO
20	2585358CD1	143	S22 S100 T71		PROTEIN AUTOIMMUNOGENIC CANCER/TESTIS ANTIGEN NYESO1 LAGE1 TRANSMEMBRANE LAGE1A LAGE1B PD091015: D7-P139	BLAST_PRODROM
21	3961495CD1	1488	S79 S152 S266 S269 S418 S424 S485 S518 S609 S614 S672 S679 S689 S729 S808 S836 S865 S905 S978 S1031 S1084 S1112 S1136 S1137 S1170 S1172 S1307 S1355 S1356 S1461	N30 N83 N275 N300 N303 N416 N425 N432 N444 N476 N540 N560 N571 N627 N685 N864 N900 N974 N1219 N1265 N1344 N1374	UBA/TTS-N domain: I737-K776	HMMER_PFAM
22	7500801CD1	505	T162 T195 T212 T259 T371 T434 T459 T509 T573 T623 T814 T1221 T1327 Y1142 S89 S270 S384 S457 T151 T203 T216 T365 T501	N177	SUBMAXILLARY APOMUCIN ICE NUCLEATION PROTEIN FILAMENTOUS HEMAGGLUTININ ANTIGEN S312 PD011940: A12-A450, M1-I519  signal_cleavage: M1-A19	BLAST_PRODROM  SPSCAN
					Signal Peptide: M1-V16, M1-A19, M1-R21, M1-P22, M1-G25	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
22					Leucine Rich Repeat: H101-R124, S77-S100, S125-D148	HMMER_PFAM
					Leucine rich repeat C-terminal domain: N158-R207	HMMER_PFAM
					Leucine zipper pattern: L129-L150	MOTIFS
23	7506414CD1	1544	S37 S67 S90 S100 S341 S423 S432 S515 S532 S547 S668 S688 S703 S835 S868 S1028 S1093 S1139 S1223 S1267 S1292 S1400 S1527 S1528	N290	Leucine Rich Repeat: N60-M82, N152-V174, A336-T358, N290-T312, L244-K266, S37-L59, A106-R128, Q267-E289, S129-A151, N198-R220, K175-P197, R221-V243, K313-V335, E359-N381, Q83-K105	HMMER_PFAM
			T400 T436 T475 T530 T607 T689 T984 T1083 T1336 T1355 T1374 T1463		PDZ domain (Also known as DHR or GLGF).: E918-D1006, T728-E814, E1014-G1105	HMMER_PFAM
					PDZ domain proteins (Als PF00595: L966-N976	BLIMPS_PFAM
					PROTEIN SH3 DOMAIN REPEAT PD00289: G969-D982	BLIMPS_PRODOM
					PROTEIN DOMAIN PROTEASE PHOSPHATASE SH3 REPEAT PDZ TYROSINE PRECURSOR HYDROLASE PD000073: E1014-P1118	BLAST_PRODOM
					GLGF DOMAIN DM00224 I38757 213-307: A720-V811 P31016 54-148: I722-V811, G912-L992	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
24	7506415CD1	1609	S37 S67 S90 S100 S341 S423 S432 S515 S532 S547 S668 S688 S703 S814 S847 S1093 S1158 S1204 S1288 S1332 S1357 S1465 S1592 S1593	N290	Leucine Rich Repeat: Q83-K105, N60-M82, N152-V174, A336-T358, N290-T312, L244-K266, S37-L59, A106-R128, Q267-E289, S129-A151, N198-R220, K175-P197, R221-V243, K313-V335, E359-N381	HMMER_PFAM
			T400 T436 T475 T530 T607 T689 T1049 T1148 T1401 T1420 T1439 T1528		PDZ domain (Also known as DHR or GLGF): E983-D1071, T707-E793, H840-E928, E1079-G1170	HMMER_PFAM
					PDZ domain proteins (Als PF00595: L888-N898	BLIMPS_PFAM
					PROTEIN SH3 DOMAIN REPEAT PD00289: G1034-D1047	BLIMPS_PRODOM
					PROTEIN DOMAIN PROTEASE PHOSPHATASE SH3 REPEAT PDZ TYROSINE PRECURSOR HYDROLASE PD000073: E1079-P1183	BLAST_PRODOM
					GLGF DOMAIN DM00224	BLAST_DOMO
					I 38757 309-402: L844-L923	
					I 31016 150-243: L844-L914	
					I 31016 54-148: E976-L1057, L708-V790, L844-L923	
					I 38757 213-307: L708-V790, E976-L1057, L844-E926	



Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
25	72192179CD1	895	S8 S17 S24 S77 S82 S107 S111 S154 S156 S225 S238 S262 S275 S279 S382 S394 S399 S459 S490 S495 S506 S523 S597 S696 S731 S736 S760 S766 S793 S818 S833 S848 S888 T58 T67 T90 T97 T127 T183 T363 T385 T393 T437 T444 T475 T483 T535 T566 T616 T655 T803 T829 T855	N69 N192 N269 N315 N624 N662 N729		
26	7505908CD1	109	S69 T64	N36 N74	Mago nashi protein: E50-I109, D4-E49 Tau and MAP proteins tubulin-binding domain signature: K41-G101 PROTEIN MAGO NASHI-LIKE NASHI MMMAGO XLMAGO R09B3.5 PD009481: D29-I109, M1-E49	HMMER_PFAM PROFILESCAN BLAST_PRODROM
27	6590147CD1	334	S80 S117 S251 S277 S316 T58 T103 T278	N302	signal_cleavage: M1-T58	SPSCAN
					Cyclin, N-terminal domain: F14-H100	HMMER_PFAM
					Cyclins proteins BL00292: L44-D74	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
27					CYCLIN I PD114253: H112-M334 CYCLIN I PD069480: M1-K33	BLAST_PRODUM BLAST_PRODUM
28	6828539CDI	569	S14 S17 S26 S59 S92 S160 S195 S201 S233 S294 S312 S341 S368 S396 S409 S422 S530 S547 T5 T54 T86 T173 T305 T387 Y316	N358 N525	EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 SH3 DOMAIN PHOSPHORYLATION PD011987: M30-L190, P247-P404	BLAST_PRODUM
29	7170321CDI	429	S5 S13 S160 S222 S276 S351 S373 T182 T187 T201 T421 Y225 Y299		CELL DIVISION PROTEIN CYCLE CDC37 HOMOLOG SIMILARITY PROTEINS K07E8.6 CONTROL PD008338: K200-E354	BLAST_PRODUM
					CELL DIVISION PROTEIN CYCLE CDC37 HOMOLOG CONTROL SIMILARITY PROTEINS CDC37LIKE PD009283: M1-K101	BLAST_PRODUM
					CELL DIVISION CYCLE PROTEIN CDC37 HOMOLOG SIMILARITY PROTEINS CONTROL PD009536: P353-P395	BLAST_PRODUM
					CDC37 HOMOLOG PROTEIN PD034382: N396-S428	BLAST_PRODUM

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
30	7505918CD1	776	S75 S112 S117 S151 S176 S211 S219 S246 S265 S364 S430 S560 S592 S614 S658 S669 S726 T51 T85 T121 T536 T697 T760 Y690	N244 N268 N381 N402 N736	PROTEIN NUCLEOPORIN INTERACTING COMPONENT NUCLEAR TRANSPORT NUCLEOPORIN AN4A AN4B DEAD EYE PD011532: M226-F480	BLAST_PRODUM
					PROTEIN T32G6.14 AN4A DEAD EYE KIAA0095 PD017605: M1-W224	BLAST_PRODUM
					PROTEIN NUCLEOPORIN INTERACTING COMPONENT NUCLEAR TRANSPORT NUCLEOPORIN AN4A KIAA0095 DEAD EYE PD014203: R481-N682, D604-E728, T605-N776	BLAST_PRODUM
31	7505935CD1	975	S15 S129 S131 S171 S266 S323 S337 S391 S412 S461 S502 S531 S564 S619 S623 S660 S685	N397 N681	CXORF5 717A PROTEIN 717A PD182566: N555- E837	BLAST_PRODUM
			S730 S731 S735 S774 S789 S811 S826 S862 S888 S906 S940 S941 S949 S957 S967		717A CXORF5 PROTEIN PD155691: M1-K199	BLAST_PRODUM
			T9 T26 T35 T107 T365 T551 T706 T802 Y342 Y353 Y437 Y751 Y879		CXORF5 717A PROTEIN 717A PD171508: F338- H425	BLAST_PRODUM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
31					PROTEIN COILED COIL CHAIN MYOSIN REPEAT HEAVY ATP-BINDING FILAMENT HEPTAD PD000002: E792-E947, E792-E961	BLAST_PRODOM
					TRICHOHYALIN DM03839  P37709 632-1103: L144-Q550, R205-E514  P22793 921-1475: E206-E556, K192-E514, Q162-Q524	BLAST_DOMO
					YEAST MYOSIN-LIKE PROTEIN MLP1 DM07884 Q02455 35-1728: S76-E931, A101-E969, Q16-N555, E792-S974	BLAST_DOMO
32	4225965CD1	814	S53 S58 S68 S101 S112 S132 S265 S285 S300 S321 S385 S613 T16 T48 T344 T472 T556 T627 T783 T793 Y172 Y591 Y673	N52 N302 N330 N340	WD domain, G-beta repeat: F486-D522, R615-D651, V570-S605, K294-N330, G194-D228, 1403-G439, L528-563	HMMER_PFAM
					WD repeats: G191-D240, R291-N330, K346-G383, T400-G439, A483-D522, T566-S605, K612-D651	HMMER_SMART
					Trp-Asp (WD) repeat protein BL00678: T511-W521	BLIMPS_BLOCKS
					G-protein beta WD-40 repeat signature: F509-G523	BLIMPS_PRINTS
					Beta G-protein (transducin) signature PR00319: P587-Y604	BLIMPS_PRINTS
					HUEMAP Echinoderm microtubule associated protein PD024545: L7-W121	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
32					Protein HUEMAP echinoderm microtubule associated F38A6.2 PD021346: N107-P231	BLAST_PRODOM
					Echinoderm microtubule-associated protein-like F38A6.2 PD182744: E449-G690	BLAST_PRODOM
					Protein F38A6.2 echinoderm microtubule associated PD040953: I238-H325	BLAST_PRODOM
					MSII, Membrane, repetitive, YDR128W, microtubule-associated protein 77k - sea urchin (Strongylocentrotus purpuratus) PM00299/A55665530-575: G610-K657	BLAST_DOMO
33	7495594CD1	3570	S31 S59 S79 S93 S128 S141 S161 S187 S210 S214 S248 S293 S356 S427 S452 S475 S484 S528	N29 N186 N347 N585 N713 N748 N847 N919 N931 N1048 N1092 N1102 N1475 N1733	signal_cleavage: M1-G17 Signal Peptide: M1-V15, M1-A19, M1-T20, M1-Q22, M1-M24	SPSCAN HMMER
			S547 S662 S669 S679 S725 S819 S835 S880 S900 S985 S1017 S1022 S1064 S1088 S1156 S1168	N1890 N1978 N2119 N2161 N2370 N2417 N2474 N2705 N2782 N2789 N2807 N2875	Domain abundant in complement control pro: C1847-C1900, C2264-C2318, C1905-C1958, C2716-C2769, C2146-C2199, C2774-C2827, C1690-C1743, C3240-C3293, C438-C493, C2497-C2550, C1789-C1842, C2948-C3001, C2832-C2885, C2555-C2607, C3415-C3467, C1963-C2016, C2890-C2943, C3298-C3351, C2381-C2435, C2083-C2141, L2653-C2711, C2323-C2376, C378-C433, C2204-C2259, C3180-C3235, C1631-C1685, C2440-C2492, C3063-C3116, C3356-C3410, C3006-C3058, C3121-C3175, C2021-C2078, C498-C559, C727-C787	HMMER_SMART

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
33			S1366 S1378 S1406 S1414 S1459 S1504 S1565 S1595 S1702 S1742 S1800 S1814 S1906 S1941	N2983 N2991 N3017 N3185 N3216 N3313 N3341 N3378 N3458 N3561	EGF-like domain: C1235-C1266, C1387-C1418, C1349-C1380, C1197-C1228, C1311-C1342, C1273-C1304, C1749-C1783, C3503-C3530, C3535-C3562, C3467-C3498, C281-L316	HMMER_PFAM
			S2010 S2020 S2052 S2092 S2109 S2121 S2184 S2203 S2307 S2348 S2509 S2741 S2763 S2766		HYR domain: I643-I722, D561-V642	HMMER_PFAM
			S2773 S2784 S2984 S2985 S3079 S3088 S3176 S3205 S3369 S3380 S3407 S3488 S3563		Epidermal growth factor-like domain: E1386-E1419, E1234-E1267, E1272-E1305, E1310-G1343, I3502-H3531, E1196-E1229, E1748-A1784, V3470-E3499, V3534-S3563, E1348-E1381, H276-T321, A322-E374, L419-Q481, F987-K1033, D1051-L1087	HMMER_SMART
			T142 T237 T327 T355 T592 T625 T690 T705 T752 T764 T849 T887 T950 T981 T1095 T1230 T1359		Calcium-binding EGF-like domain: N1345-E1381, N1383-E1419, D1231-E1267, N1269-E1305, E1307-G1343, D1745-A1784, E1196-E1229, C3503-H3531, C3471-E3499, C3535-S3563	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
33			T1397 T1643 T1828 T1870 T1980 T1988 T2185 T2219 T2403 T2464 T2491 T2536 T2594 T2720		Pentraxin / C-reactive protein / pentaxin: T1420-F1627	HMMER SMART
			T2809 T2813 T2857 T3019 T3104 T3155 T3296 T3300 T3568 Y388 Y804 Y858 Y1465 Y2045 Y2632		von Willebrand factor (vWF) type A domain: R81-L260	HMMER SMART
					Pentaxin family: S1470-W1608	HMMER PFAM
					Sushi domain (SCR repeat): C2146-C2199, C1690-C1743, C2264-C2318, C1789-C1842, C2497-C2550, C2832-C2885, C1847-C1900, C438-C493, C2774-C2827, C1905-C1958, C1963-C2016, C2948-C3001, C2555-C2607, C3240-C3293, C3415-C3467, C2381-C2435, C2890-C2943, C3121-C3175, C3063-C3116, C1631-C1685, C3298-C3351, T2659-C2711, C2204-C2259, C2716-C2769, C3356-C3410, C2083-C2141, C2323-C2376, C3006-C3058, C2021-C2078, C378-C433, C2440-C2492, C3180-C3235, C498-C559, C727-C787	HMMER PFAM
					von Willebrand factor type A domain: E83-A259	HMMER PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
33					Pentaxin family proteins BL00289: G1425-F1455, H1512-G1530, L1551-V1589, G1603-G1617	BLIMPS_BLOCKS
					Calcium-binding EGF-like domain proteins pattern proteins BL01187: C1322-F1337	BLIMPS_BLOCKS
					Pentaxin signature PR00895: L1448-D1462, H1512-G1530, L1539-D1558, K1559-N1578, N1578-L1592	BLIMPS_PRINTS
					Type II EGF-like signature PR00010: N1269-N1280, N1403-F1413, S1414-T1420	BLIMPS_PRINTS
					Calcium-binding precursor PD00919: C1349-C1360	BLIMPS_PRODROM
					PRECURSOR SIGNAL PENTAXIN PROTEIN GLYCOPROTEIN PLASMA C-REACTIVE CALCIUM ACUTE PHASE PD002153: Y1438-W1608	BLAST_PRODROM
					GLYCOPROTEIN THYROGLOBULIN PRECURSOR REPEAT THYROID HORMONE IODINATION SIGNAL EGF-LIKE PROTEIN PD009765: T1014-C1160, G991-C1160, T1068-C1235	BLAST_PRODROM
					SIMILARITY TO EGF-TYPE REPEATS GLYCOPROTEIN EGF-LIKE DOMAIN PD147765: W550-L844	BLAST_PRODROM



Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
33					SUSHI REPEAT DM04887 P27113 I-551: D1732-I2093, C2769-S3078, T2678-F3021, K1843-Y2162, Y2848-A3132, P2069-S2459, S3205-P3538, S2348-P2615, C2518-T2901, Q3030-I3362, E3082-V3421, C360-D572, C281-I567, P265-N551, W2369-T2624, D1581-G1793, C1387-E1419, C1273-F1329, C451-A584, C1197-S1276, V721-C787, C988-T1041, G698-Y757	BLAST_DOMO
					SUSHI REPEAT DM04887 P1658 I-609: D1732-N2133, Y1805-P2207, T2346-P2615, N2673-K2944, W2067-P2446, Y2856-G3134, G3150-P3538, T2669-P2893, L3040-E3411, C277-P570, S2407-C2612, H1701-S1867, T1594-C1743, C1273-E1431, C1235-L1428, C1197-S1276, C999-D1129, K723-C787	BLAST_DOMO
					SUSHI REPEAT DM04887 P33730 I-610: Y1805-P2207, D1732-N2133, I2714-C3006, F2102-E2439, C2172-G2568, T2346-P2615, Y2848-C3121, I3173-P3538, Q2931-I3362, S2671-K2944, G292-D572, G2654-G2903, S266-N551, S1595-Q1816, E2470-F2633, T1594-C1743, C1235-F1426, C1197-F1337, N712-W776, R3468-W3557, V721-C787, P696-T779, Q1078-G1107	BLAST_DOMO
					C-REACTIVE PROTEIN DM00835 P4797 I194-431: F1426-G1617	BLAST_DOMO
					Aspartic acid and asparagine hydroxylation site: C1246-C1257, C1284-C1295, C1322-C1333, C1360-C1371, C1398-C1409, C1761-C1772	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
33					EGF-like domain signature 1: C1217-C1228, C1255-C1266, C1293-C1304, C1331-C1342, C1369-C1380, C1407-C1418, C3487-C3498, C3519-C3530, C3551-C3562	MOTIFS
					EGF-like domain signature 2: C305-C320, C360-C373, C1217-C1228, C1255-C1266, C1293-C1304, C1331-C1342, C1369-C1380, C1407-C1418, C1770-C1783, C3487-C3498, C3519-C3530	MOTIFS
					Calcium-binding EGF-like domain pattern signature: D1231-C1255, N1269-C1293, E1307-C1331, N1345-C1369, N1383-C1407, D1745-C1770	MOTIFS

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
34/7722608CB1/ 3696	1-3312, 105-747, 176-287, 176-325, 176-326, 192-456, 253-325, 261-549, 272-575, 371-1015, 398-995, 615-914, 616-905, 616-1040, 623-1000, 667-973, 731-1349, 751-1402, 846-1522, 848-1443, 850-1130, 850-1187, 888-1178, 966-1419, 1047-1259, 1079-1230, 1079-1234, 1100-1230, 1123-1600, 1264-1671, 1292-1618, 1580-1687, 1585-1687, 2645-3354, 2702-3411, 2718-3224, 2721-3410, 2763-3024, 2763-3268, 2774-3527, 2778-3299, 2784-3608, 2804-3316, 2819-3396, 2828-3299, 2840-3454, 2852-3442, 2882-3611, 2885-3236, 2913-3663, 2953-3242, 2960-3658, 2963-3416, 2976-3651, 3003-3284, 3010-3652, 3012-3280, 3020-3562, 3022-3659, 3025-3290, 3031-3633, 3040-3649, 3079-3446, 3085-3346, 3086-3269, 3086-3696, 3107-3675, 3108-3619, 3113-3358, 3113-3477, 3135-3431, 3136-3366, 3151-3474, 3151-3481, 3151-3670, 3153-3501, 3155-3403, 3157-3397, 3176-3456, 3180-3396, 3181-3659, 3189-3384, 3189-3419, 3189-3673, 3189-3674, 3202-3453, 3202-3616, 3209-3266, 3215-3495, 3215-3656, 3216-3459, 3219-3251, 3219-3252, 3242-3660
35/7505869CB1/ 5585	1-501, 1-580, 1-5576, 5-505, 5-569, 22-590, 34-600, 34-818, 87-531, 87-705, 115-648, 137-342, 157-690, 157-769, 221-428, 224-516, 235-714, 235-787, 295-798, 545-773, 549-1048, 714-1302, 757-1176, 771-1267, 776-1308, 782-986, 795-1296, 813-1052, 837-1217, 960-1211, 989-1527, 990-1243, 1033-1260, 1103-1393, 1147-1429, 1156-1344, 1156-1416, 1164-1593, 1167-1946, 1183-1394, 1183-1458, 1209-1512, 1209-1695, 1254-1726, 1277-1521, 1367-1844, 1568-2037, 1568-2072, 1593-2184, 1669-1809, 1669-2297, 1762-2314, 1820-2454, 1880-2390, 1883-2116, 1883-2355, 1883-2389, 1884-2086, 1909-2157, 1951-2270, 1951-2461, 1963-2218, 1996-2609, 2015-2308, 2023-2287, 2047-2291, 2084-2551, 2111-2541, 2252-2471, 2265-2922, 2272-2497, 2300-2929, 2331-2594, 2365-2614, 2365-2633, 2365-2855, 2389-2644, 2389-2783, 2395-2985, 2421-2711, 2421-2943, 2421-2968, 2446-2693, 2469-2938, 2469-2971, 2482-2931, 2520-2792, 2521-2758, 2521-2800, 2521-3072, 2538-2785, 2543-2962, 2549-2911,
	2555-2967, 2568-2774, 2570-2787, 2570-2805, 2570-2854, 2570-2857, 2570-2881, 2570-2882, 2570-2888, 2575-2986, 2602-2985, 2604-2982, 2673-3065, 2732-2985, 2753-2981, 2786-3072, 2820-2986, 2925-3294, 2926-3170, 3012-3344, 3013-3245, 3070-3292, 3071-3327, 3090-3331, 3094-3350, 3129-3329, 3189-3326, 3204-3468, 3225-3812, 3238-3907, 3314-3501, 3317-3584, 3348-3608, 3420-3680, 3421-3675, 3442-3715, 3453-3639, 3465-3759, 3519-3829, 3532-4092, 3537-3999, 3537-4001, 3570-4012, 3581-3823, 3585-4177, 3585-4179, 3590-4067, 3635-4010, 3707-4078, 3715-3986, 3715-4200, 3721-4012, 3738-4007, 3794-4006, 3839-4357, 3849-4411, 3853-4414, 3904-4404, 3934-4010, 3947-4584, 3954-4526, 3966-4193, 3998-4193, 3998-4265, 3998-4398, 4020-4634, 4033-4393, 4062-4694, 4071-4630, 4078-4526, 4121-4689, 4149-4794, 4199-4788, 4209-4825, 4217-4505, 4217-4637, 4217-4666, 4217-4691, 4217-4701, 4217-4736, 4217-4746, 4217-4765, 4217-4776, 4217-4778, 4224-4626,

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
	4264-4759, 4281-4788, 4285-4864, 4297-4817, 4322-4564, 4331-4540, 4331-4630, 4370-4763, 4371-4941, 4444-4829, 4449-4630, 4480-4947, 4490-4580, 4498-4628, 4517-4728, 4559-5193, 4561-4920, 4592-4737, 4594-5083, 4607-4874, 4617-4910, 4617-4911, 4620-5190, 4625-5303, 4637-5199, 4653-4893, 4655-4900, 4655-5138, 4672-5191, 4676-5247, 4677-5187, 4700-5311, 4703-4962, 4708-4817, 4713-4920, 4714-5207, 4722-5295, 4728-4990, 4742-5051, 4745-5270, 4776-5075, 4805-5352, 4850-5444, 4858-5153, 4879-5492, 4893-5245, 4917-5553, 4944-5555, 4946-5556, 4955-5436, 4989-5558, 4996-5408, 5032-5565, 5045-5277, 5065-5580, 5070-5318, 5070-5408, 5070-5549, 5073-5336, 5081-5336, 5081-5364, 5099-5566, 5102-5566, 5106-5313, 5106-5585, 5108-5570, 5112-5559, 5116-5577, 5118-5583, 5121-5565, 5121-5566, 5122-5566, 5123-5563, 5123-5566, 5132-5567, 5148-5572, 5155-5567, 5155-5571, 5174-5567, 5183-5566, 5190-5567, 5201-5584, 5204-5455, 5242-5524, 5263-5566, 5351-5544, 5376-5521, 5438-5566
36/7502420CB1/ 2667	1-570, 284-777, 285-1010, 298-855, 317-774, 317-930, 318-792, 318-843, 318-852, 319-558, 319-691, 319-858, 319-924, 319-952, 319-980, 319-982, 319-990, 319-1028, 319-1029, 319-1154, 323-1042, 323-1592, 325-620, 330-829, 333-850, 333-905, 345-763, 345-1071, 346-602, 347-906, 350-1007, 360-988, 370-636, 371-1071, 373-967, 377-653, 384-919, 385-607, 385-919, 386-924, 414-1147, 486-641, 505-736, 505-1071, 606-1157, 816-1354, 819-1284, 828-1359, 828-1373, 832-1419, 848-1568, 851-1566, 853-1566, 856-1569, 858-1564, 865-1579, 896-1582, 915-1579, 1001-1548, 1002-1570, 1009-1548, 1104-1336, 1125-1592, 1145-1500, 1167-1726, 1203-1427, 1225-1849, 1239-1568, 1261-1540, 1261-1948, 1262-1594, 1277-1937, 1288-1781, 1311-1796, 1375-1605, 1377-1555, 1377-1591, 1426-1974, 1457-1893, 1532-1785, 1553-2140, 1562-2041, 1587-2213, 1588-1858, 1599-2116, 1603-1850, 1685-1848, 1685-2281, 1701-2314, 1709-2286, 1710-1976, 1712-2261, 1763-2382, 1817-2469, 1843-2464, 1854-2102, 1867-2443, 1868-2430, 1870-2525, 1871-2491, 1883-2472, 1888-2667, 1902-2485, 1918-2184
37/7506455CB1/ 1559	1-260, 1-272, 147-288, 147-382, 147-402, 147-403, 147-686, 151-427, 152-430, 155-425, 155-431, 159-341, 164-413, 411-667, 500-775, 545-1152, 558-858, 701-1218, 701-1252, 715-1184, 840-1235, 868-1214, 877-1142, 879-1225, 884-1199, 916-1342, 933-1194, 947-1231, 970-1223, 980-1198, 980-1226, 1012-1225, 1034-1226, 1041-1224, 1046-1231, 1052-1559, 1072-1382
38/7506018CB1/ 645	1-172, 1-645, 27-172, 27-304, 32-172, 40-167, 54-156, 62-172, 84-401, 84-645, 137-582, 172-645, 173-324, 173-363, 173-366, 173-369, 173-390, 173-392, 173-397, 173-398, 173-402, 173-403, 173-417, 173-424, 173-426, 173-428, 173-429, 173-432, 173-433, 173-439, 173-447, 173-449, 173-452, 173-453, 173-482, 173-490, 173-534, 173-590, 173-628, 173-629, 173-630, 173-633, 173-634, 173-637, 173-641, 173-645, 174-629, 174-644, 175-413, 177-434, 177-435, 177-466, 177-578, 178-629, 179-423, 179-634, 179-645, 181-484, 181-542, 182-461, 182-630, 183-619, 183-640, 185-433, 185-442, 185-629, 185-636, 185-645, 186-639, 186-645, 188-626, 192-438, 192-449, 192-454, 193-492, 194-497, 195-645, 196-378, 197-506, 197-634, 198-637, 199-477, 200-471, 200-525, 200-631, 201-480, 202-389, 202-636, 204-445, 204-453, 204-466, 204-495, 204-562, 204-636, 204-644, 207-587, 207-639, 208-630, 208-635, 209-629, 210-645, 213-630, 213-631, 213-634, 214-519, 214-634, 215-645, 216-629, 217-473, 217-489, 218-451, 218-501,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	218-633, 218-645, 219-433, 219-632, 220-494, 220-506, 220-634, 221-637, 222-628, 222-633, 223-479, 223-542, 223-630, 223-641, 223-645, 224-472, 224-485, 224-631, 226-634, 228-633, 230-435, 232-636, 233-454, 237-517, 239-370, 239-506, 240-638, 241-487, 241-636, 241-645, 242-633, 244-478, 244-629, 245-527, 251-521, 253-564, 253-577, 253-645, 258-636, 260-523, 263-517, 263-628, 266-645, 269-508, 269-521, 270-645, 271-445, 272-440, 272-504, 272-535, 273-634, 274-640, 275-630, 277-631, 280-634, 281-634, 282-634, 282-636, 282-645, 283-634, 284-632, 285-631, 285-634, 288-636, 289-482, 289-519, 289-549, 289-636, 289-645, 290-636, 291-536, 291-544, 291-627, 292-634, 293-533, 293-577, 296-566, 297-631, 298-575, 298-644, 299-587, 299-636, 299-643, 301-576, 305-636, 307-590, 307-593, 307-629, 307-636, 309-642, 310-628, 311-521, 311-561, 312-555, 312-638, 313-597, 314-557, 315-628, 316-629, 316-636, 321-563, 321-597, 321-634, 321-636, 322-634, 324-633, 332-459, 332-635, 338-536, 341-606.
	341-626, 343-625, 343-629, 344-608, 347-562, 347-631, 348-553, 348-572, 349-636, 350-645, 352-610, 352-622, 353-465, 353-585, 353-586, 358-612, 358-634, 360-595, 360-606, 360-634, 361-604, 362-636, 363-573, 363-597, 364-489, 364-590, 364-639, 365-609, 366-625, 368-562, 368-645, 372-510, 372-585, 373-598, 373-621, 381-634, 383-631, 386-636, 388-636, 391-636, 392-522, 392-629, 393-614, 393-642, 395-636, 399-620, 399-629, 399-645, 401-645, 402-597, 402-631, 402-645, 404-645, 405-636, 407-491, 407-598, 407-645, 411-629, 411-645, 412-632, 412-645, 413-645, 414-643, 416-634, 419-633.
	420-633, 420-645, 423-610, 423-631, 424-645, 425-629, 426-638, 428-632, 429-634, 429-645, 430-628, 430-630, 431-631, 432-634, 433-631, 436-636, 437-634, 439-635, 441-634, 441-645, 444-645, 446-595, 446-630, 446-636, 446-645, 449-636, 449-645, 453-634, 456-631, 458-545, 464-641, 465-645, 467-645, 469-630, 471-566, 471-636, 472-629, 472-632, 472-645, 478-645, 479-636, 491-636, 492-640, 494-634, 494-645, 529-639, 531-645, 543-644, 560-645, 572-636.
39/353017CB1/ 1824	1-645, 1-1824, 287-825, 567-1155, 738-1346, 816-875, 818-1311, 828-875, 875-1081, 875-1291, 1101-1651, 1101-1653, 1101-1654, 1164-1223, 1168-1824, 1184-1212
40/5137816CB1 8043	1-721, 41-696, 105-695, 114-353, 114-484, 114-659, 114-713, 176-434, 176-712, 176-724, 176-744, 176-747, 176-817, 178-855, 208-534, 296-855, 325-1048, 343-987, 349-855, 391-941, 442-625, 496-667, 596-1158, 635-978, 669-1138, 685-1298, 697-1169, 738-1335, 746-1357, 769-1359, 830-1463, 835-1322, 846-1383, 857-997, 861-1523, 882-1519, 914-1424, 936-1601, 938-1193, 947-1467, 952-1477, 952-1633, 952-1653, 955-1237, 970-1606, 1011-1716, 1024-1175, 1024-1589, 1024-1592, 1043-1628, 1060-1222, 1078-1589, 1112-1778, 1147-1790, 1163-1790, 1167-1790, 1178-1790, 1181-1408, 1186-1790, 1203-1743.

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1220-1790, 1241-1863, 1242-1850, 1248-1796, 1254-1790, 1257-1556, 1261-1834, 1261-1861, 1281-1445, 1307-1790, 1345-1724, 1440-2053, 1453-1790, 1476-1950, 1493-1882, 1525-2169, 1543-2022, 1575-1790, 1588-1790, 1711-2299, 1726-2334, 1914-2097, 1914-2319, 1914-2321, 1933-2438, 2050-2350, 2102-2598, 2102-2690, 2111-2418, 2136-2301, 2230-2885, 2409-2922, 2414-3075, 2440-3073, 2596-3136, 2596-3138, 2920-3410, 2966-8043, 2979-3344, 3111-3410, 3230-3400, 3424-4037, 3530-3558, 3587-3610, 3589-3610, 3639-3898, 3724-4558, 3732-4444, 3870-4014, 3878-4558
41/7502151CBI/ 1550	1-687, 151-767, 162-862, 162-873, 172-841, 179-811, 192-1001, 207-820, 604-1050, 604-1178, 633-1023, 649-1188, 685-1291, 761-1301, 817-1297, 818-1301, 848-1332, 852-1251, 897-1439, 903-1301, 921-1550
42/7505983CBI/ 2108	1-158, 1-224, 1-330, 1-440, 1-2108, 3-335, 3-435, 3-436, 4-320, 13-215, 14-279, 15-304, 18-270, 20-305, 21-211, 30-271, 39-312, 48-393, 85-329, 97-390, 124-414, 178-369, 232-440, 332-776, 332-846, 332-861, 332-890, 332-958, 437-1097, 459-1104, 491-697, 508-765, 536-1249, 548-1134, 604-1098, 619-1221, 620-972, 627-847, 680-980, 680-1098, 680-1184, 680-1239, 680-1262, 680-1283, 680-1499, 714-976, 715-1258, 724-1321, 732-1187, 737-1025, 737-1403, 751-1378, 763-1031, 767-1403, 782-991, 787-1444, 800-1272, 826-1321, 827-1227, 829-1430, 832-1353, 866-1413, 874-1042, 879-1155, 882-1430, 894-1112, 896-1529, 905-1352, 910-1337, 938-1521, 946-1451, 951-1216, 955-1586, 956-1556, 961-1230,
	962-1606, 981-1683, 984-1223, 989-1576, 999-1581, 1007-1379, 1007-1380, 1010-1573, 1014-1692, 1021-1106, 1032-1445, 1044-1603, 1052-1286, 1056-1482, 1069-1313, 1083-1350, 1086-1287, 1086-1581, 1104-1362, 1104-1427, 1107-1353, 1117-1391, 1124-1366, 1130-1717, 1136-1797, 1142-1667, 1152-1325, 1152-1391, 1152-1394, 1152-1420, 1152-1421, 1153-1405, 1158-1413, 1158-1662, 1161-1787, 1177-1375, 1184-1373, 1199-1765, 1206-1733, 1229-1880, 1243-1501, 1245-1849, 1253-1558,
	1257-1483, 1259-1791, 1261-1524, 1264-1633, 1284-1546, 1284-1808, 1288-1748, 1288-1873, 1288-1907, 1293-1971, 1297-1580, 1312-1524, 1315-1906, 1318-2044, 1323-1909, 1325-1741, 1329-1787, 1332-1581, 1333-1581, 1333-1676, 1333-1992, 1339-2101, 1340-1998, 1340-2035, 1341-1614, 1342-1814, 1354-1637, 1354-2031, 1356-1985, 1361-2105, 1362-2048, 1364-1948, 1377-1645, 1382-1594, 1383-1636, 1385-1675, 1385-1681, 1419-1887, 1420-2032, 1421-2073, 1422-2094, 1439-1742, 1443-2108, 1451-1746, 1452-2097, 1457-2031, 1458-1751, 1460-2097, 1463-2085, 1467-2089, 1471-2087, 1472-2108, 1473-2108, 1479-1899, 1483-1993, 1500-2108, 1507-2108, 1508-2099, 1510-2037, 1510-2101, 1514-2101, 1517-1764, 1518-1932, 1518-1934, 1531-2091, 1533-1780, 1533-1801, 1533-2044, 1540-1686, 1542-2102, 1543-1960, 1543-2101, 1546-1756, 1546-2075, 1546-2108, 1556-1687, 1558-1694, 1565-2103, 1567-1887, 1586-2108, 1589-2102, 1599-1910, 1602-2059, 1615-1812, 1618-1896,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1619-1933, 1619-2108, 1623-1902, 1624-2102, 1632-2108, 1635-1896, 1637-2089, 1639-2102, 1639-2104, 1645-2103, 1646-2097, 1647-2104, 1649-2104, 1650-2102, 1652-2082, 1654-2108, 1659-2104, 1663-2104, 1665-2065, 1665-2106, 1668-2102, 1668-2108, 1669-2108, 1671-2096, 1673-2102, 1677-1854, 1678-2101, 1679-2102, 1689-2104, 1690-2107, 1692-2108, 1693-2104, 1694-2097, 1695-2104, 1704-2108, 1708-1935, 1715-1964, 1717-1970, 1717-1972, 1717-1989, 1718-2103, 1725-2103, 1728-2104, 1733-1977, 1738-1908, 1738-2096, 1739-1976, 1739-2102, 1741-2102, 1752-2034, 1755-1993, 1755-2104, 1762-2108, 1764-2108, 1775-2104, 1777-2102, 1782-2102, 1782-2104, 1783-2105, 1786-2104, 1788-2101, 1789-2085, 1796-2104, 1796-2108, 1804-2105, 1805-2108, 1808-2104, 1810-2037, 1810-2102, 1812-2049, 1833-2101, 1856-2102, 1862-2108, 1871-2104, 1872-2101, 1872-2107, 1881-2108, 1905-2102, 1910-2107, 1936-2104, 1942-2104, 1954-2104, 1955-2102, 1957-2108, 1987-2101, 1989-2108, 2044-2108
43/7505986CB1/ 524	1-158, 1-224, 1-330, 1-509, 3-335, 3-410, 4-320, 13-215, 14-279, 15-304, 18-270, 20-305, 21-211, 30-271, 39-312, 47-499, 48-393, 67-499, 85-329, 97-390, 124-410, 178-369, 232-410, 439-524
44/3231075CB1/ 1987	1-1987, 276-552, 276-951, 316-585, 397-658, 400-728, 408-986, 448-1069, 449-1181, 467-949, 467-1064, 528-1216, 736-1006, 736-1156, 736-1231, 824-1225, 844-1233, 847-1047, 1079-1256
45/7503516CB1/ 2857	1-2692, 554-767, 587-1181, 588-1121, 898-1017, 1765-2329, 1813-2362, 1821-2126, 1821-2479, 1827-2106, 1827-2113, 1831-2425, 1854-2453, 1857-2209, 1863-2130, 1868-2104, 1882-2351, 1889-2016, 1896-2471, 1904-2144, 1905-2560, 1922-2590, 1929-2191, 1936-2235, 1943-2199, 1963-2573, 1969-2252, 1970-2205, 1985-2206, 1985-2573, 1996-2168, 1997-2316, 2005-2573, 2014-2697, 2015-2694, 2023-2692, 2026-2217, 2042-2558, 2042-2692, 2047-2579, 2053-2345, 2059-2692, 2065-2299, 2070-2203, 2091-2609, 2104-2484, 2107-2692, 2107-2701, 2114-2394, 2115-2299, 2117-2409, 2133-2400, 2136-2692, 2136-2410, 2143-2643, 2143-2857, 2145-2857, 2148-2692, 2149-2578, 2152-2431, 2153-2694, 2156-2520, 2159-2514, 2165-2692, 2166-2692, 2172-2599, 2173-2655, 2175-2653, 2181-2641, 2181-2692, 2184-2314, 2184-2642, 2184-2688, 2184-2692, 2184-2695, 2184-2699, 2185-2692, 2185-2694, 2186-2693, 2187-2692, 2188-2690, 2188-2692, 2192-2692, 2195-2689, 2197-2434, 2206-2670, 2207-2648, 2211-2641, 2211-2700, 2214-2641, 2217-2634, 2218-2655, 2222-2682, 2223-2692, 2223-2698, 2231-2692, 2254-2700, 2255-2688, 2255-2692, 2255-2699, 2262-2401, 2264-2691, 2268-2689, 2269-2654, 2269-2692, 2272-2641, 2272-2653, 2273-2654, 2274-2670, 2275-2692, 2276-2692, 2297-2692, 2299-2692, 2300-2694, 2301-2692, 2306-2696, 2308-2691, 2308-2695, 2314-2641, 2329-2692, 2332-2689, 2334-2692, 2340-2586, 2340-2692, 2346-2655, 2353-2682, 2356-2641, 2360-2602, 2360-2692, 2361-2692, 2365-2692, 2372-2692, 2378-2675, 2378-2677, 2388-2641, 2402-2692, 2403-2692, 2411-2567, 2430-2641, 2436-2692, 2442-2579, 2498-2692, 2617-2692

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
46/7506179CB1/ 2375	1-387, 9-300, 15-279, 15-393, 15-2373, 27-393, 28-292, 31-393, 32-269, 33-297, 35-253, 35-276, 35-300, 35-466, 35-577, 36-527, 39-464, 39-728, 40-597, 41-297, 41-305, 43-305, 43-334, 43-386, 44-372, 63-258, 64-336, 73-386, 76-373, 78-323, 113-805, 150-386, 150-681, 151-434, 155-756, 159-458, 161-545, 167-689, 176-404, 183-757, 184-534, 201-368, 204-372, 205-375, 214-667, 218-474, 252-454, 261-769, 269-513, 269-550, 269-829, 269-942, 304-901, 314-576, 336-843, 362-610, 370-659, 373-819, 377-823, 382-526, 384-618, 398-668, 404-942, 406-988, 443-759, 443-985, 492-786, 496-609, 523-794, 543-1198, 555-805, 561-914, 565-973, 581-988, 609-844, 611-907, 639-787, 644-1329, 644-1361, 658-1280, 661-809, 661-1219, 675-1225, 689-1126, 699-792, 704-1359, 715-983, 716-825, 738-1296, 765-1228, 781-1357, 787-1431, 794-1378, 814-1215, 847-1491, 859-1236, 903-1290, 989-1159, 989-1244, 989-1256, 989-1269, 989-1581, 995-1329, 1001-1480, 1007-1615, 1009-1237, 1014-1191, 1016-1253, 1024-1286, 1034-1616, 1038-1314, 1047-1191, 1052-1325, 1078-1465, 1088-1629, 1101-1765, 1106-1557, 1108-1331, 1110-1768, 1111-1887, 1119-1464, 1126-1433, 1132-1347, 1134-1304, 1134-1350, 1134-1543, 1134-1568, 1134-1600, 1134-1601, 1134-1604, 1134-1624, 1134-1644, 1134-1645, 1134-1658, 1134-1662, 1134-1669, 1134-1677, 1134-1678, 1134-1701, 1134-1712, 1137-1592, 1142-1791, 1151-1429, 1152-1967, 1155-1423, 1159-1482, 1165-1807, 1169-1841, 1173-1397, 1188-1634, 1190-1617, 1190-1841, 1192-1580, 1192-1840, 1194-1461, 1195-1406, 1203-1804, 1205-1731, 1207-1838, 1217-1824, 1224-1817, 1226-1863, 1240-1826, 1242-1883, 1255-1781, 1266-1875, 1276-1546, 1277-1514, 1288-1583, 1292-1775, 1310-1652, 1317-1762, 1320-1563, 1320-1800, 1322-1662, 1341-1403, 1341-1645, 1345-2177, 1347-1568, 1347-1735, 1347-1736, 1347-1754, 1356-1522, 1366-1623, 1370-1616, 1376-1830, 1381-1770, 1390-1688, 1393-1729, 1400-1624, 1400-1669, 1400-1999, 1406-1579, 1422-2059, 1439-2046, 1442-1979, 1452-1649, 1452-1982, 1454-1546, 1458-1649, 1458-2049, 1466-1753, 1466-1773, 1467-1903, 1485-1772, 1487-2078, 1488-1878, 1490-1790, 1497-1826, 1500-1926, 1500-2141, 1512-2016, 1518-1884, 1518-2073, 1524-2349, 1525-1860, 1529-1784, 1538-1711, 1538-1838, 1542-2217, 1542-2227, 1542-2229, 1543-1775, 1543-1806, 1549-1875, 1550-1992, 1551-2114, 1554-1712, 1555-1748, 1560-1705, 1560-2034, 1570-1815, 1570-1826, 1578-1824, 1604-2175, 1604-2268, 1608-2073, 1608-2085, 1609-1758, 1614-1934, 1627-1848, 1627-1933, 1627-2059, 1628-2344, 1632-1912, 1637-1757, 1637-2322, 1647-2021, 1652-1918, 1658-2303, 1661-2341, 1663-1772, 1663-1927, 1663-1931, 1664-1920, 1664-1972, 1668-2282, 1676-2307, 1681-1934, 1683-2072, 1692-2353, 1694-1947, 1694-1972, 1703-2196, 1709-1958, 1714-2041, 1722-1975, 1722-2001, 1725-1985, 1732-2186, 1734-2181, 1743-2032, 1748-2196, 1748-2357, 1750-2194, 1752-2189, 1773-2001, 1777-2010, 1779-2022,



Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
47/1938744CB1/ 486	1779-2189, 1799-2227, 1801-2353, 1806-1929, 1813-2357, 1817-2060, 1819-2121, 1822-2270, 1824-2195, 1834-2186, 1835-2087, 1835-2375, 1844-2085, 1844-2113, 1846-2076, 1846-2173, 1847-2196, 1856-2132, 1857-2111, 1858-2195, 1862-2106, 1865-2353, 1867-2353, 1883-2026, 1884-2121, 1884-2242, 1884-2338, 1886-2147, 1887-2195, 1888-2375, 1890-2374, 1897-2189, 1897-2375, 1904-2352, 1906-2105, 1909-2358, 1913-2374, 1920-2353, 1924-2375, 1925-2337, 1927-2189, 1932-2230, 1938-2259, 1938-2352, 1939-2351, 1940-2353, 1942-2352, 1942-2358, 1943-2246, 1943-2353, 1943-2375, 1945-2352, 1947-2351, 1948-2371, 1959-2352, 1959-2372, 1961-2353, 1977-2353, 1978-2276, 1983-2351, 1984-2353, 1988-2352, 1990-2351, 1994-2196, 1998-2231, 2003-2209, 2003-2352, 2008-2261, 2015-2352, 2015-2353, 2017-2236, 2018-2374, 2020-2289, 2034-2272, 2056-2352, 2079-2316, 2084-2358, 2085-2358, 2088-2196, 2088-2353, 2092-2347, 2092-2358, 2108-2358, 2115-2353, 2117-2353, 2120-2283, 2132-2358, 2142-2351, 2142-2353, 2145-2353, 2208-2359
48/5557436CB1/ 1527	1-187, 1-1495, 38-607, 134-888, 165-755, 169-736, 301-751, 356-743, 399-1032, 399-1054, 433-1030, 501-772, 501-819, 512-652, 522-741, 527-634, 534-788, 577-819, 667-1324, 684-974, 692-819, 710-1117, 777-1212, 803-1068, 841-874, 841-888, 841-917, 841-921, 841-935, 841-989, 841-1007, 841-1042, 841-1052, 841-1061, 841-1070, 841-1077, 841-1090, 841-1121, 841-1131, 841-1201, 841-1421, 841-1500, 853-1521, 854-941, 855-1504, 856-940, 859-1496, 861-1527, 894-1024, 896-1513, 910-1147, 923-1159, 930-1200, 935-1196, 935-1512, 944-1512, 950-1168, 950-1191, 950-1284, 950-1373, 950-1410, 950-1415, 950-1436, 950-1467, 950-1474, 950-1517, 957-1512, 958-1527, 964-1515, 978-1527, 994-1279, 997-1273, 1040-1527, 1041-1314, 1448-1494
49/7506178CB1/ 1422	1-147, 1-509, 1-654, 3-245, 3-285, 3-428, 3-467, 3-481, 3-1422, 47-525, 76-549, 96-362, 208-451, 208-531, 210-491, 211-714, 220-472, 223-812, 228-476, 229-550, 229-808, 230-358, 230-450, 230-504, 230-526, 232-516, 233-491, 233-523, 234-537, 236-714, 237-514, 237-525, 242-508, 242-509, 242-537, 242-592, 243-489, 244-511, 244-769, 247-518, 247-528, 249-478, 249-499, 249-512, 252-503, 253-833, 255-523, 255-531, 255-801, 256-505, 273-632, 329-557, 337-590, 337-833, 340-627, 349-633, 368-596, 386-622, 388-666, 534-693, 550-798, 550-831, 551-825, 555-836, 581-814, 604-636, 615-869, 736-1040, 837-1055, 837-1082, 837-1095, 837-1102, 837-1110, 837-1322, 837-1399, 838-1375, 842-1097, 853-1090,
	853-1095, 856-1119, 857-1113, 858-1087, 865-1129, 874-1136, 882-1111, 898-1125, 899-1180, 933-1406, 938-1422, 940-1406, 954-1411, 957-1184, 957-1397, 957-1409, 957-1413, 958-1410, 959-1410, 960-1406, 961-1411, 962-1406, 963-1406, 963-1411, 964-1405, 964-1409, 965-1406, 966-1413, 970-1411, 972-1238, 972-1240, 976-1401, 979-1422, 983-1261, 984-1262, 984-1275, 989-1410, 995-1162, 995-1406, 1002-1266, 1009-1406, 1010-1233, 1017-1410, 1023-1408, 1027-1406, 1034-1248, 1040-1192, 1072-1410, 1076-1411, 1079-1345, 1107-1410, 1135-1406, 1137-1400, 1139-1406, 1145-1395, 1153-1379, 1183-1410, 1186-1422, 1188-1422, 1224-1421

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
50/7506235CB1/ 1822	1-339, 1-1630, 5-261, 7-281, 7-431, 93-341, 126-327, 126-386, 126-427, 126-430, 196-431, 230-457, 429-682, 429-876, 429-1062, 429-1111, 439-715, 442-600, 443-652, 455-937, 455-1137, 521-1151, 534-1300, 566-1249, 589-1283, 593-1180, 609-1180, 609-1296, 616-1211, 620-1245, 634-1167, 670-1168, 689-1174, 696-1066, 700-1365, 706-1167, 719-1245, 722-1454, 794-1492, 806-1105, 839-1104, 842-1267, 849-1528, 867-1444, 872-1006, 873-1084, 873-1504, 897-1317, 920-1481, 930-1632, 960-1233, 960-1309, 1079-1535, 1098-1378, 1123-1392, 1125-1757, 1127-1387, 1166-1491, 1180-1493, 1194-1807, 1196-1544, 1210-1455, 1223-1479, 1234-1822, 1241-1679, 1245-1809, 1262-1424, 1262-1482, 1276-1482, 1276-1493, 1281-1809, 1283-1807, 1284-1491, 1291-1482, 1298-1482, 1336-1630, 1338-1614, 1490-1579, 1620-1649, 1752-1781
51/1302184CB1/ 6104	1-455, 15-545, 62-568, 62-586, 147-442, 147-573, 164-614, 240-673, 345-6088, 476-1023, 494-1148, 1156-1761, 1165-1729, 1294-1552, 1295-1782, 1399-1749, 1402-1663, 1467-1734, 1550-2076, 1551-2024, 1897-2416, 1906-2521, 2008-2304, 2191-2596, 2320-2621, 2335-2625, 2428-2591, 2466-3109, 2659-3152, 2713-3437, 2761-3475, 2763-2940, 2817-3531, 2821-3052, 2972-3108, 3028-3416, 3154-3465, 3156-3535, 3166-3416, 3193-3465, 3282-3803, 3477-4087, 3509-3865, 3723-4076, 3905-4809, 3920-4162, 4387-4597, 4557-4700, 4768-5103, 4812-5089, 4823-5171, 4823-5319, 4823-5334, 4823-5357, 4823-5427, 4823-5432, 4823-5437, 4823-5466, 4823-5481, 4823-5622, 4829-5014, 4829-5536, 4948-5124, 5010-5720, 5019-5313, 5099-5245, 5147-6038, 5171-5451, 5179-5889, 5214-5372, 5214-5547, 5214-5696, 5214-5740, 5214-5792, 5214-5793, 5240-5544, 5276-5787, 5346-6032, 5356-6046, 5368-5965, 5429-5771, 5462-6010, 5465-5760, 5474-5720, 5481-6069, 5512-5791, 5522-5959, 5524-6103, 5524-6104, 5560-6072, 5563-6032, 5642-6088, 5644-6089, 5656-6087, 5664-6104, 5685-6090, 5790-6088, 5831-6075
52/7506232CB1/ 1137	1-279, 4-251, 4-269, 4-286, 4-309, 4-458, 7-287, 8-261, 9-306, 10-231, 10-268, 10-294, 11-280, 11-313, 11-567, 16-235, 17-262, 18-352, 19-269, 20-496, 20-771, 21-300, 21-497, 22-292, 23-290, 24-311, 25-283, 29-265, 29-266, 29-282, 29-323, 30-321, 33-497, 37-326, 37-362, 48-497, 55-294, 66-259, 69-325, 69-359, 71-497, 76-331, 90-497, 99-436, 117-378, 119-497, 127-343, 131-322, 138-731, 143-412, 155-406, 162-497, 174-395, 212-498, 213-351, 218-480, 255-392, 255-497, 256-541, 286-497, 316-594, 332-497, 341-497, 493-746, 493-790, 494-1137, 495-771, 496-724, 503-777
53/2585358CB1/ 1011	1-296, 31-240, 36-752, 70-558, 71-558, 86-549, 89-356, 112-915, 113-553, 122-417, 123-557, 191-540, 225-558, 364-975, 364-978, 367-988, 380-846, 394-780, 399-977, 422-960, 427-734, 427-843, 508-907, 519-972, 550-958, 550-1009, 551-850, 551-906, 551-933, 551-954, 551-955, 551-961, 551-962, 553-905, 553-954, 553-972, 555-957, 556-914, 556-934, 556-1003, 559-899, 559-967, 559-971, 559-982, 559-986, 559-990, 559-992, 559-993, 559-1006, 559-1007, 559-1011, 560-1003, 565-991, 566-977, 569-907, 569-958, 569-960, 569-971, 569-977, 575-838, 575-918, 575-977, 580-821, 580-991, 583-993, 584-802, 584-951, 588-993, 612-991, 621-892, 666-952, 686-951, 698-993, 709-993, 755-963, 765-977

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
54/3961495CB1/ 4638	1-640, 6-339, 66-547, 66-4638, 69-560, 100-730, 212-4638, 284-904, 390-775, 657-900, 747-1019, 747-1125, 781-1542, 945-1625, 948-1526, 1130-1752, 1157-2019, 1190-1736, 1313-1691, 1363-1676, 1570-1750, 1570-1871, 1573-2371, 1583-1723, 1583-1827, 1583-1858, 1583-1859, 1583-2235, 1583-2250, 1583-2321, 1583-2340, 1583-2342, 1583-2388, 1618-2061, 1796-2087, 1796-2088, 1850-2361, 1925-2729, 1939-2206, 1940-2130, 1968-2395, 1976-2729, 1979-2729, 1985-2242, 1985-2536, 1986-2262, 2088-2729, 2095-2729, 2123-2345, 2123-2473, 2142-2453, 2151-2423, 2153-2536, 2172-2487, 2224-2981, 2261-2536, 2262-2536, 2329-2536, 2334-2874, 2337-2981, 2397-2536, 2450-2536, 2454-2765, 2454-2989, 2544-3214, 2627-2925, 2753-3175, 2818-3040, 2831-3449, 2854-3101, 2919-3722, 3306-3783, 3884-4635, 3911-4638, 3946-4635, 3967-4635, 4021-4638, 4127-4638, 4137-4635, 4138-4633, 4154-4638, 4166-4638, 4170-4638, 4179-4638, 4189-4631, 4204-4638, 4223-4638, 4228-4613, 4292-4638, 4369-4638, 4566-4638
55/7500801CB1/ 2233	1-604, 1-2233, 139-561, 322-600, 322-1004, 322-1062, 322-1075, 328-831, 328-912, 331-601, 331-1094, 351-635, 365-610, 375-595, 375-900, 377-781, 377-992, 377-996, 406-949, 415-558, 427-679, 498-1074, 528-809, 534-784, 544-1105, 551-788, 554-767, 565-792, 575-841, 577-1028, 581-1095, 582-845, 583-1012, 591-1200, 596-930, 609-952, 609-1158, 611-1154, 637-917, 651-1150, 654-1100, 663-1050, 682-850, 692-1071, 696-1013, 702-944, 713-1159, 725-973, 736-991, 744-1265, 749-907, 763-1144, 832-1206, 848-1455, 871-990, 879-1303, 903-1394, 928-1314, 948-1235, 967-1203, 975-1117, 1012-1597, 1015-1396, 1024-1186, 1029-1420, 1030-1321, 1034-1316, 1053-1234, 1078-1722, 1107-1316, 1118-1342, 1120-1426, 1126-1380, 1126-1449, 1151-1381, 1178-1399, 1200-1330, 1212-1595, 1216-1343, 1223-1651, 1233-1632, 1414-1629, 1472-1823, 1499-1646, 1532-1661, 1536-1863, 1579-1834, 1597-1756, 1597-1802, 1597-1886, 1630-1873, 1663-1873, 1677-1894, 1686-2233, 1688-1894, 1741-1874, 1800-1872, 1800-1873, 1894-1983, 1894-2127, 1894-2194, 1894-2195, 1894-2196, 1894-2223, 1894-2230, 1894-2233, 1896-2182, 1896-2233, 1897-2175, 1897-2233, 1900-2143, 1900-2233, 1901-2233, 1902-2195, 1905-2233, 1912-2233, 1918-2216, 1918-2218, 1928-2182, 1942-2233, 1943-2233, 1951-2108, 1970-2182, 1976-2233, 1982-2120, 2038-2233, 2158-2233
56/7506414CB1/ 4884	1-4882, 399-983, 512-738, 522-601, 523-601, 528-1162, 629-1159, 684-1374, 760-1418, 914-1466, 914-1508, 922-1513, 928-1513, 929-1498, 935-1513, 944-1584, 944-1642, 971-1434, 993-1513, 996-1507, 999-1508, 1000-1714, 1005-1507, 1015-1507, 1017-1507, 1041-1513, 1052-1507, 1080-1507, 1082-1672, 1082-1733, 1132-2010, 1287-1821, 1330-2013, 1398-2028, 1404-2016, 1421-1524, 1434-2016, 1440-2016, 1461-1600, 1467-1945, 1483-2093, 1487-2016, 1504-2016, 1528-2016, 1551-2016, 1557-2016, 1584-2016, 1607-2019, 1621-2197, 1653-2520, 1852-2088, 1994-2613, 2045-2721, 2047-2354, 2099-2276, 2101-2334, 2120-2546, 2155-2741, 2178-2494, 2207-2252, 2312-2877, 2471-3373, 2609-2654,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	2752-2998, 2752-3183, 2752-3299, 2752-3347, 2775-3347, 2805-3072, 2806-3060, 2806-3630, 2821-3341, 2822-3275, 2861-3154, 2861-3725, 2882-3394, 2905-3530, 2907-3150, 2928-3440, 2970-3495, 2979-3392, 2980-3495, 3075-3350, 3113-3682, 3128-3643, 3140-3309, 3179-3681, 3239-3419, 3304-3860, 3364-3630, 3462-3559, 3463-3735, 3519-4170, 3528-3820, 3530-3820, 3533-3822, 3536-4124, 3548-3947, 3566-3947, 3594-3997, 3633-4337, 3694-4062, 3711-3969, 3763-4041, 3795-4181, 3851-4381, 3860-4119, 3868-4105, 3868-4116, 3877-4409, 3881-3958, 3895-4147, 3934-4178, 3940-4826, 3946-4074, 3961-4430, 3976-4379, 4026-4645, 4082-4113, 4089-4178, 4089-4628, 4118-4632, 4152-4872, 4177-4388, 4180-4379, 4209-4388, 4213-4510, 4214-4465, 4242-4497, 4242-4503, 4394-4701, 4407-4720, 4418-4726, 4418-4872, 4421-4870, 4432-4870, 4437-4860, 4437-4870, 4437-4872, 4437-4882, 4442-4872, 4444-4882, 4456-4872, 4457-4872, 4459-4873, 4460-4872, 4463-4704, 4464-4873, 4465-4872, 4468-4732, 4473-4872, 4478-4717, 4485-4842, 4489-4871, 4489-4872, 4490-4872, 4507-4742, 4522-4872, 4533-4872, 4534-4872, 4534-4876, 4535-4872, 4536-4872, 4539-4875, 4541-4873, 4546-4848, 4546-4875, 4556-4872, 4557-4882, 4578-4849, 4584-4873, 4589-4872, 4594-4872, 4599-4872, 4601-4855, 4606-4854, 4610-4783, 4610-4874, 4611-4872, 4617-4835, 4617-4875, 4638-4765, 4647-4780, 4651-4877, 4654-4872, 4659-4868, 4660-4875, 4662-4882, 4670-4873, 4681-4878, 4698-4875, 4702-4879, 4712-4875, 4714-4872, 4722-4882, 4726-4882, 4731-4872, 4735-4872, 4742-4882, 4742-4884, 4743-4884, 4744-4882, 4750-4882, 4771-4870, 4781-4882, 4792-4875
5177506415CB1/ 5079	1-5077, 399-983, 512-738, 522-601, 528-601, 528-1162, 629-1159, 684-1374, 760-1418, 914-1466, 914-1508, 922-1513, 928-1513, 929-1498, 935-1513, 944-1584, 944-1642, 971-1434, 993-1513, 996-1507, 999-1508, 1000-1714, 1005-1507, 1015-1507, 1017-1507, 1041-1513, 1052-1507, 1080-1507, 1082-1672, 1082-1733, 1132-2010, 1287-1821, 1330-2013, 1398-2028, 1404-2016, 1420-2220, 1421-1524, 1434-2016, 1440-2016, 1461-1600, 1467-1945, 1483-2093, 1487-2016, 1504-2016, 1520-2312, 1528-2016, 1551-2016, 1557-2016, 1584-2016, 1607-2019, 1616-2435, 1811-2287, 1852-2088, 2079-2618, 2114-2678, 2115-2431, 2144-2189, 2163-2747, 2546-2591, 2719-3462, 2733-2864, 2842-3196, 2925-3542, 2930-3559, 2940-3494, 2944-3378, 2945-3193, 2970-3542, 3000-3267, 3001-3255, 3001-3825, 3016-3536, 3017-3470, 3056-3349, 3056-3920, 3077-3589, 3100-3725, 3102-3345, 3123-3635, 3165-3690, 3174-3587, 3175-3690, 3270-3545, 3308-3877, 3323-3838, 3335-3504, 3374-3876, 3434-3614, 3499-4055, 3559-3825, 3657-3754, 3658-3930, 3714-4365, 3723-4015, 3725-4015, 3728-4017, 3731-4319, 3743-4142, 3756-4142, 3789-4192, 3828-4532, 3889-4257, 3906-4164, 3958-4236, 3990-4376, 4046-4576, 4055-4314, 4063-4300, 4063-4311, 4072-4604, 4076-4153, 4090-4342, 4129-4373, 4135-5021,

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
	4141-4269, 4156-4625, 4171-4574, 4221-4840, 4277-4308, 4284-4373, 4284-4823, 4313-4827, 4347-5067, 4372-4583, 4375-4574, 4404-4583, 4408-4705, 4409-4660, 4437-4692, 4437-4698, 4589-4896, 4602-4915, 4613-4921, 4613-5067, 4616-5065, 4627-5065, 4632-5055, 4632-5065, 4632-5067, 4632-5077, 4637-5067, 4639-5077, 4651-5067, 4652-5067, 4654-5068, 4655-5067, 4658-4899, 4659-5068, 4660-5067, 4663-4927, 4668-5067, 4673-4912, 4680-5037, 4684-5066, 4684-5067, 4685-5067, 4702-4937, 4717-5067, 4728-5067, 4729-5067, 4729-5071, 4730-5067, 4731-5067, 4734-5070, 4736-5068,
	4741-5043, 4741-5070, 4751-5067, 4752-5077, 4773-5044, 4779-5068, 4784-5067, 4789-5067, 4794-5067, 4796-5050, 4801-5049, 4805-4978, 4805-5069, 4806-5067, 4812-5030, 4812-5070, 4833-4960, 4842-4975, 4846-5072, 4849-5067, 4854-5063, 4855-5070, 4857-5077, 4865-5068, 4876-5073, 4893-5070, 4897-5074, 4907-5070, 4909-5067, 4917-5077, 4921-5077, 4926-5067, 4930-5067, 4937-5077, 4937-5079, 4938-5079, 4939-5077, 4945-5077, 4966-5065, 4976-5077, 4987-5070
58/72192179CB1/ 6065	1-814, 556-1091, 580-908, 598-1091, 599-739, 610-1079, 618-1086, 749-1421, 1026-1395, 1120-1729, 1183-1920, 1184-1807, 1248-1479, 1248-1497, 1248-1529, 1248-1624, 1261-1490, 1261-1499, 1296-1814, 1335-2012, 1354-1630, 1381-1615, 1381-1732, 1381-1765, 1402-1920, 1444-2140, 1473-2154, 1605-2278, 1807-2420, 1858-2394, 1868-2278, 1937-2449, 1985-2513, 2030-2457, 2082-2513, 2161-2869, 2165-2428, 2169-2711, 2239-2814, 2257-2421, 2257-2485, 2257-2493, 2261-2438, 2323-3003, 2436-2674, 2505-3051, 2554-2778, 2554-2780, 2711-3364, 2714-3077, 2758-3045, 2871-3128, 2874-3473, 2899-3159, 2899-3166, 2945-3159, 2949-3580, 2967-3113, 3026-3649, 3036-3632, 3071-3896, 3091-3521, 3091-3633, 3134-3350, 3174-3782, 3277-3918, 3280-3521, 3280-3625, 3282-3909, 3295-3990, 3320-3959, 3325-3587, 3325-3777, 3363-3902, 3421-3507, 3466-3926, 3476-4061, 3480-4135,
	3489-4107, 3531-4107, 3534-4187, 3550-4016, 3554-4209, 3564-4162, 3587-3803, 3603-3760, 3624-3799, 3626-4221, 3674-4049, 3695-3820, 3695-4124, 3727-4365, 3778-3992, 3781-4421, 3829-4440, 3880-4414, 3910-4154, 3930-4603, 3962-4126, 3996-4255, 3996-4559, 4004-4296, 4039-4371, 4049-4392, 4049-4600, 4079-4256, 4079-4264, 4079-4685, 4080-4563, 4094-4771, 4105-4521, 4106-4724, 4159-4410, 4170-4746, 4177-4561, 4178-4416, 4194-4690, 4204-4685, 4213-4667, 4227-4687, 4236-4690, 4255-4852, 4260-4809, 4277-4584, 4287-4494, 4294-4687, 4306-4553, 4306-4729, 4317-4908, 4339-4622, 4367-4961, 4417-4644, 4449-4926, 4460-4696, 4465-4824, 4475-4584, 4486-4759, 4591-4740,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	4626-4850, 4632-4878, 4663-4897, 4668-4885, 4698-5421, 4713-5352, 4747-4931, 4762-5580, 4763-5015, 4781-5041, 4786-5422, 4791-5430, 4795-4938, 4830-5088, 4846-5649, 4856-5047, 4861-5052, 4904-5122, 4906-5670, 4909-5138, 4909-5151, 4922-5624, 4936-5242, 4995-5539, 4996-5278, 5001-5257, 5035-5335, 5055-5733, 5058-5332, 5086-5733, 5087-5720, 5088-5340, 5094-5684, 5098-5375, 5111-5631, 5121-5383, 5132-5778, 5157-5687, 5165-5383, 5169-5813, 5178-5566, 5187-5779, 5192-5447, 5198-5518, 5198-5624, 5285-5543, 5311-5571, 5311-5898, 5333-5573, 5333-5576, 5361-5614, 5389-5607, 5417-5693, 5444-6021, 5446-6023, 5453-5672, 5453-5735, 5479-6024, 5481-6057, 5488-5675, 5496-6023, 5509-6059, 5516-6061, 5525-6065, 5527-5876, 5528-6059, 5540-5798, 5552-6018, 5558-6059, 5592-6059, 5612-5885, 5615-5999, 5637-6065, 5641-6061, 5643-5908, 5644-5933, 5645-5858, 5657-6065, 5665-5938, 5680-5812, 5697-5938, 5738-6037, 5752-6056, 5752-6065, 5780-6056, 5784-6065, 5799-6061, 5853-6063, 5913-6041, 5913-6065
59/7505908CB1/ 1055	1-297, 5-261, 7-283, 14-274, 21-180, 21-267, 34-203, 34-296, 34-307, 34-522, 43-280, 211-504, 211-535, 215-526, 218-514, 245-512, 272-512, 280-503, 293-498, 315-415, 378-570, 426-550, 443-1055
60/6590147CB1/ 1651	1-666, 1-672, 21-655, 296-815, 296-971, 322-1093, 324-897, 330-942, 360-1236, 369-1393, 372-1083, 380-986, 380-1085, 390-1020, 401-944, 411-1060, 411-1061, 412-843, 419-1230, 423-890, 423-1050, 444-1028, 448-1061, 449-1090, 449-1127, 450-942, 458-970, 465-1008, 467-1319, 468-991, 473-881, 473-1106, 473-1335, 474-779, 474-1237, 481-776, 481-797, 481-1153, 483-756, 483-884, 488-884, 488-952, 488-967, 488-988, 490-636, 491-952, 494-927, 494-1348, 495-761, 495-764, 495-830, 496-797, 497-721, 497-763, 498-997, 502-894, 506-738, 508-855, 508-888, 509-747, 512-895, 514-732, 516-858, 519-919, 522-878, 522-996, 525-770, 525-788, 525-1264, 526-801, 526-894, 527-1271, 528-818, 529-816, 532-1200, 533-763, 536-948, 536-1153, 539-982, 556-1238, 654-1381, 698-1322, 717-1395, 719-1309, 721-1236, 726-1381, 727-1381, 737-1381, 750-1381, 756-1341, 763-1261, 773-1394, 779-1373, 795-1399, 803-1372, 804-1387, 810-1320, 824-1399, 873-1274, 886-1399, 887-1396, 891-1387, 894-1399,
	897-1370, 905-1343, 906-1381, 911-1380, 912-1303, 912-1382, 913-1378, 913-1382, 913-1384, 914-1378, 914-1651, 916-1384, 917-1309, 917-1381, 919-1376, 919-1381, 920-1380, 920-1381, 920-1389, 921-1374, 921-1378, 921-1386, 922-1381, 923-1380, 923-1381, 925-1345, 925-1372, 926-1378, 927-1380, 928-1382, 928-1384, 928-1500, 931-1380, 932-1378, 933-1378, 933-1381, 934-1380, 935-1380, 935-1381, 936-1381, 937-1381, 938-1389, 939-1384, 939-1406, 939-1459, 940-1384, 941-1381, 942-1381, 942-1382, 944-1382, 944-1389, 945-1381, 945-1382, 945-1432, 947-1374, 948-1381, 948-1427, 949-1384, 950-1372, 950-1374, 950-1378, 950-1381, 950-1383, 950-1439, 950-1454, 953-1399, 954-1374, 954-1395, 954-1508, 955-1385, 955-1387, 956-1380, 958-1381, 959-1381, 960-1381, 962-1378, 963-1380, 964-1380, 966-1379, 966-1381, 966-1454, 968-1399, 969-1378, 970-1381

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
61/6828539CB1/ 2015	1-152, 1-256, 1-287, 1-384, 1-409, 1-596, 1-612, 1-790, 2-432, 6-410, 6-480, 11-445, 121-747, 122-916, 140-390, 168-380, 260-692, 412-924, 417-903, 417-926, 420-926, 425-897, 425-926, 437-882, 438-905, 441-926, 445-908, 449-926, 455-926, 457-918, 469-926, 478-908, 479-951, 480-918, 489-918, 497-908, 498-926, 504-885, 504-926, 506-903, 509-917, 526-926, 556-914, 560-908, 594-926, 614-924, 616-926, 622-926, 641-926, 645-918, 655-1330, 659-883, 661-924, 671-926, 690-892, 855-1592, 918-1167, 951-1406, 995-1677, 1072-1277, 1146-1339, 1163-1405, 1168-1671, 1257-2015
62/7170321CB1/ 1760	1-160, 1-296, 4-235, 12-154, 28-158, 28-742, 42-312, 92-668, 92-768, 209-1755, 422-688, 422-894, 432-952, 542-1024, 577-1184, 578-1368, 580-1187, 584-1387, 592-1220, 599-1291, 600-1275, 607-1142, 614-1185, 620-1172, 625-1344, 632-1392, 633-1300, 637-1305, 639-1197, 651-1243, 652-1303, 664-1287, 671-1319, 674-1250, 674-1506, 677-1181, 677-1266, 691-1347, 712-955, 712-1328, 714-1380, 715-1338, 725-969, 729-1345, 731-961, 731-1344, 731-1395, 734-1255, 744-1347, 750-1387, 756-1305, 757-1344, 758-1341, 766-1354, 771-1344, 775-1341, 785-1459, 800-1041, 808-1022, 810-1416, 825-1404, 838-1371, 861-1082, 864-1438, 866-1111, 866-1135, 872-1295, 884-1422, 889-1315, 889-1527, 899-1545, 901-1308, 940-1723, 943-1755, 944-1460, 959-1643, 972-1624, 1008-1653, 1014-1588, 1024-1557, 1029-1468, 1047-1641, 1048-1591, 1055-1583, 1060-1755, 1063-1570, 1065-1732, 1072-1585, 1073-1471, 1078-1659, 1084-1587, 1093-1755, 1100-1682, 1107-1681, 1107-1703, 1132-1760, 1134-1715, 1139-1755, 1144-1655, 1150-1702, 1153-1760, 1154-1635, 1154-1745, 1156-1565, 1156-1713, 1158-1758, 1158-1760, 1161-1717, 1165-1760, 1186-1735, 1194-1759, 1197-1718, 1232-1754, 1236-1755, 1240-1756, 1241-1760, 1251-1753, 1251-1760, 1255-1755, 1263-1692, 1263-1753, 1264-1758, 1265-1755, 1268-1695, 1268-1760, 1270-1755, 1279-1753, 1280-1760, 1290-1760, 1294-1734, 1294-1756, 1299-1760, 1301-1753, 1305-1707, 1305-1760, 1306-1753, 1307-1757, 1311-1760, 1314-1760, 1316-1760, 1317-1755, 1322-1714, 1333-1754, 1333-1755, 1334-1753, 1335-1753, 1335-1755, 1336-1755, 1338-1755, 1341-1753, 1342-1753, 1343-1753, 1345-1753, 1347-1755, 1348-1753, 1349-1760, 1352-1760, 1358-1755, 1359-1760, 1362-1755, 1367-1755, 1368-1755, 1369-1758, 1371-1755, 1372-1753, 1373-1760, 1375-1753, 1376-1758
63/7505918CB1/ 2599	1-2599, 867-1311, 1386-1645, 1389-1645, 1620-1953, 2170-2579, 2230-2567, 2230-2594, 2381-2599
64/7505935CB1/ 3305	1-311, 1-488, 7-3305, 10-294, 159-414, 159-459, 188-793, 206-811, 253-424, 254-708, 266-813, 328-835, 353-803, 373-817, 388-835, 416-635, 416-716, 416-790, 428-835, 432-835, 505-826, 616-1455, 642-826, 717-843, 750-958, 801-1058, 850-1097, 869-1134, 964-1250, 1118-1722, 1141-1753, 1181-1761, 1225-1467, 1225-1681, 1254-1433, 1271-1527, 1277-1517, 1323-1598, 1330-1615, 1373-1978, 1373-2048, 1383-1594, 1411-1665, 1415-2070, 1416-2034, 1435-1647, 1435-1869, 1452-2069, 1452-2107, 1454-1710, 1461-1608, 1496-1996, 1502-1755, 1502-1849, 1509-1762, 1510-2158, 1548-2148, 1553-2164, 1575-2114, 1581-2018, 1620-2274, 1636-1971, 1639-2139, 1648-2100, 1651-1913, 1653-1939,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1664-1776, 1685-1953, 1695-2255, 1721-2071, 1737-2586, 1749-2261, 1751-1956, 1751-1970, 1751-2205, 1778-2436, 1806-2391, 1808-2357, 1825-2318, 1864-2113, 1877-2100, 1877-2307, 1884-2150, 1893-2403, 1894-2485, 1897-2392, 1900-2531, 1905-2508, 1905-2547, 1907-2606, 1921-2561, 1924-2590, 1926-2590, 1942-2505, 1959-2551, 1977-2134, 1988-2523, 2008-2528, 2011-2613, 2017-2558, 2020-2334, 2022-2304, 2022-2334, 2052-2613, 2057-2595, 2069-2446, 2069-2611, 2081-2613, 2094-2365, 2095-2378, 2105-2613, 2108-2346, 2112-2766, 2124-2381, 2131-2438, 2185-2445, 2201-2450, 2205-2325, 2208-2420, 2271-2540, 2314-2643, 2315-2553, 2315-2613, 2318-2583, 2321-2613, 2343-2640, 2362-2613, 2368-2602, 2368-2635, 2459-3148, 2520-2783, 2520-2998, 2595-2888, 2598-3169, 2614-2776, 2614-2891, 2614-2894, 2614-3158, 2615-2884, 2622-2823, 2622-2827, 2665-2912, 2669-2929, 2676-2765, 2676-2803, 2691-2826, 2691-3169, 2698-3165, 2702-2924, 2704-3169, 2711-3169, 2712-3165, 2714-3169, 2717-3161, 2721-2987, 2723-3165, 2723-3167, 2724-3169, 2727-3165, 2727-3167, 2732-3014, 2743-3126, 2751-3167, 2769-3169, 2771-3168, 2772-3165, 2773-3167, 2774-3167, 2780-3169, 2800-3165, 2810-3077, 2810-3166, 2810-3167, 2815-3169, 2828-3168, 2830-3167, 2833-3167, 2837-3296, 2839-3298, 2844-3298, 2845-3137, 2850-3098, 2851-3169, 2871-3167, 2873-3167, 2883-3298, 2884-3080, 2884-3112, 2885-3155, 2885-3157, 2885-3169, 2889-3298, 2891-3153, 2899-3168, 2899-3169, 2899-3305, 2907-3167, 2909-3167, 2921-3167, 2933-3169, 2941-3167, 2958-3298, 2986-3169, 2999-3165, 3018-3298, 3030-3169, 3110-3298, 3238-3305
65/4225965CBI/ 2923	1-888, 171-947, 622-976, 695-1285, 709-1360, 724-1171, 766-1151, 770-1415, 823-1534, 851-1550, 937-1238, 948-1552, 948-1590, 948-1632, 950-1511, 950-1600, 963-1653, 1006-1071, 1129-1676, 1129-1759, 1129-1906, 1181-1418, 1244-1965, 1264-1959, 1307-1606, 1310-1769, 1349-1606, 1401-2025, 1418-1881, 1418-2249, 1436-1694, 1436-1884, 1452-2046, 1464-2088, 1466-2215, 1567-2322, 1568-2056, 1569-2119, 1591-1838, 1591-2214, 1597-1808, 1597-1824, 1660-2260, 1736-2013, 1736-2014, 1736-2186, 1775-2419, 1779-2012, 1779-2036, 1788-2432, 1804-2070, 1829-2203, 1876-2422, 1900-2182, 1900-2245, 1952-2223, 1989-2219, 1989-2339, 1997-2185, 2007-2224, 2061-2324, 2117-2362, 2117-2432, 2125-2375, 2125-2649, 2161-2372, 2191-2664, 2216-2518, 2437-2634, 2465-2714, 2478-2876, 2482-2867, 2483-2875, 2490-2697, 2490-2722, 2512-2869, 2571-2811, 2649-2874, 2675-2923, 2724-2871, 2724-2876
66/7495594CBI/ 11546	1-278, 1-610, 1-1261, 564-828, 576-1089, 849-1240, 876-1046, 886-1013, 886-1163, 886-1283, 886-1379, 886-1390, 886-1403, 886-1407, 886-1418, 886-1419, 886-1424, 886-1425, 886-1436, 886-1492, 886-1507, 886-1557, 886-1601, 894-1541, 900-1654, 939-1791, 944-1232, 944-1519, 944-1524, 944-1586, 944-1600, 948-1577, 973-1600, 987-1551, 1017-1529, 1017-1666, 1066-1247, 1071-1805, 1101-1779, 1153-1666, 1171-1652, 1224-1531, 1236-1928, 1252-1946, 1256-1600, 1267-1887, 1272-1651, 1328-1698, 1348-1801, 1348-1818, 1359-2249, 1366-1671, 1366-2035, 1366-2095, 1367-1589, 1367-1637, 1367-1727, 1367-1881, 1389-2102, 1391-2060, 1399-2095, 1420-1885, 1447-2186, 1447-6998, 1463-2103, 1470-1984, 1470-2164, 1470-2188, 1471-2208, 1472-2305, 1490-2173, 1500-2212, 1503-2131, 1504-1997, 1530-2383, 1552-2254, 1555-2333, 1562-2363, 1580-2252, 1580-2302, 1593-2362,



Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
66	<p>1595-2333, 1595-2391, 1595-2403, 1628-2380, 1639-2383, 1651-2351, 1684-2331, 1694-2383, 1703-2381, 1711-2376, 1714-2383, 1721-2356, 1735-2383, 1739-2383, 1775-2383, 1794-2383, 1825-2383, 1826-2383, 1866-2383, 1875-2112, 1882-2625, 1883-2451, 1884-2513, 1892-2383, 1893-2383, 1899-2383, 1923-2392, 1930-2383, 1934-2274, 1949-2383, 1987-2383, 2006-2383, 2027-2303, 2039-2383, 2044-2383, 2059-2511, 2060-2383, 2086-2383, 2096-2892, 2098-2357, 2098-2544, 2098-2787, 2262-2760, 2314-2768, 2347-2896, 2379-2896, 2426-2749, 2436-2888, 2455-2861, 2460-2876, 2468-2896, 2474-2896, 2482-2968, 2486-2894, 2515-2896, 2536-2896, 2554-2896, 2588-2896, 2692-2880, 2692-3225, 2737-3152, 2868-3176, 2874-3174, 3070-3430, 3124-3520, 3257-3583, 3434-3926, 3434-3961, 3434-4143, 3434-4205, 3434-4281, 3436-4205, 3481-4226, 3490-4080, 3547-3982, 3554-3647, 3757-4165, 3906-4237, 4094-4508, 4197-4284, 4344-4479, 4404-4558, 4559-4589, 4559-4614, 4559-4651, 4559-4678, 4559-4737, 4559-4759, 4559-4780, 4559-4781, 4559-4871, 4559-4896, 4559-4910, 4559-4919, 4559-4936, 4561-4631, 4697-4936, 4859-5240, 4971-5572, 4971-5575, 4971-5577, 4971-5579, 4971-5580, 4972-5580, 4973-5580, 5072-5691, 5178-5580, 5179-5579, 5179-5580, 5243-5627, 5288-5650, 5373-5832, 5416-5911, 5494-5911, 5511-6034, 5656-6216, 5656-6224, 5656-6253, 5656-6283, 5656-6320, 5656-6402, 5656-6428, 5656-6498, 5656-6549, 5656-6564, 5656-6570, 5777-6589, 5864-6585, 5930-6209, 5978-6561, 6002-6653, 6058-6452, 6064-6936, 6078-6936, 6114-6545, 6174-6507, 6174-6779, 6174-6905, 6174-6970, 6174-6991, 6174-6994, 6174-6997, 6174-6999, 6175-6912, 6175-6922, 6175-6986, 6177-6936, 6181-6507, 6181-6789, 6217-6936, 6221-6999, 6229-6995, 6231-6936, 6238-6476, 6238-6999, 6255-6999, 6256-6999, 6264-6999, 6266-6999, 6267-6860, 6270-6999, 6273-6611, 6287-6999, 6290-6999, 6306-6936, 6307-6999, 6310-6999, 6327-6978, 6364-6999, 6367-6999, 6370-6936, 6385-6999, 6402-6999, 6478-6957, 6478-7115, 6527-6936, 6590-7094, 6708-6952, 6708-7259, 6714-6954, 6770-7327, 6802-7443, 6841-7260, 6855-7446, 6905-7039, 6910-7546, 6910-10894, 7598-8225, 8125-8316, 8125-8745, 8490-9296, 8767-9603, 8768-9039, 8768-9143, 8768-9395, 8768-9448, 8768-9449, 8770-9552, 8773-9269, 8775-9638, 8868-9106, 8868-9240, 8873-9724,</p> <p>9017-9908, 9034-9749, 9044-9745, 9083-9678, 9158-9819, 9174-9436, 9174-9437, 9174-9950, 9244-9479, 9244-9521, 9247-10070, 9253-9828, 9292-9793, 9340-9804, 9396-10205, 9418-10185, 9423-10142, 9425-10050, 9438-10220, 9484-10342, 9518-10263, 9518-10325, 9518-10346, 9624-10233, 9637-9884, 9637-10247, 9649-10471, 9662-10252, 9662-10275, 9666-10245, 9756-10397, 9765-10220, 9765-10348, 9796-10382, 9803-10524, 9809-10142, 9813-10620, 9815-10524, 9829-10092, 9854-10452, 9873-10530, 9876-10120, 9876-10130, 9876-10363, 9877-10088, 9879-10228, 9896-10569, 9929-10311, 9930-10186, 9938-10373, 9946-10532, 9948-10167, 9949-10694, 9987-10598, 10012-10339, 10020-10647, 10043-10276, 10043-10607, 10058-10333, 10065-10468, 10071-10350, 10154-10742, 10168-10699, 10180-11031, 10184-11022, 10190-10765, 10191-10859, 10202-10957, 10203-10428, 10203-10527, 10203-10531, 10203-10648, 10203-10735, 10203-10778, 10203-10814, 10203-10878, 10203-10942, 10203-10944, 10203-10945, 10203-11112, 10220-11003, 10224-10493, 10226-10663, 10229-10456,</p>

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
66	10229-10735, 10229-10753, 10243-10734, 10250-11048, 10251-11101, 10252-11100, 10261-10872, 10263-10463, 10263-10655, 10264-11005, 10285-10861, 10325-10604, 10325-10651, 10329-10951, 10357-11228, 10360-10783, 10361-11244, 10366-10665, 10376-10603, 10377-11246, 10379-11108, 10381-11251, 10384-10786, 10386-10697, 10387-10985, 10388-11257, 10393-10985, 10419-11263, 10421-10947, 10422-11280, 10423-10725, 10427-10881, 10429-11034, 10438-10716, 10453-10782, 10463-10677, 10464-10721, 10474-10731, 10478-11212, 10485-11027, 10492-11217, 10498-10666, 10502-10772, 10504-11260, 10510-11094, 10521-11004, 10523-11223, 10528-10646, 10528-10815, 10528-10833, 10528-10880, 10528-10892, 10528-11051, 10529-11000, 10530-11019, 10532-10737, 10532-10749, 10532-10909, 10532-10913, 10534-11266, 10540-11247, 10542-11250, 10543-11259, 10549-11111, 10551-10810, 10552-11112, 10566-11133, 10579-10881, 10580-11066, 10580-11266, 10584-10821, 10586-10815, 10602-10935, 10604-11142, 10610-11268, 10624-11013, 10634-11254, 10634-11274, 10635-11252, 10659-11274, 10678-10926, 10683-11257, 10690-11371, 10694-11223, 10721-11256, 10721-11274, 10723-11229, 10723-11274, 10726-11270, 10728-11296, 10728-11369, 10741-11368, 10760-11368, 10769-11360, 10779-11274, 10784-11274, 10799-11268, 10820-11274, 10821-11271, 10825-11366, 10829-11270, 10829-11370, 10831-11366, 10836-11269, 10841-11262, 10841-11268, 10873-11273, 10875-11360, 10880-11340, 10884-11328, 10887-11268, 10900-11372, 10901-11371, 10904-11370, 10909-11339, 10926-11371, 10935-11350, 10936-11256, 10940-11374, 10941-11368, 10955-11371, 10963-11371, 10976-11281, 10982-11352, 10982-11371, 10983-11240, 10985-11368, 10986-11268, 10988-11248, 10989-11277, 10994-11266, 11011-11336, 11013-11269, 11015-11231, 11021-11268, 11030-11260, 11031-11374, 11035-11274, 11037-11371, 11060-11270, 11075-11265, 11075-11369, 11097-11268, 11098-11546, 11103-11371, 11110-11317, 11117-11270, 11118-11331, 11148-11216, 11149-11371, 11174-11413, 11174-11441, 11184-11260, 11227-11374

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
34	7722608CB1	EOSIHET02
35	7505869CB1	FIBRTXS07
36	7502420CB1	LATRTUT02
37	7506455CB1	NOSEDIT02
38	7506018CB1	EOSIHET02
39	353017CB1	MUSLTDR02
40	5137816CB1	PLACFER01
41	7502151CB1	CERVNOT01
42	7505983CB1	LIVRNON08
43	7505986CB1	LUNGTUT17
44	3231075CB1	BMARTXT02
45	7503516CB1	HEAONOR01
46	7506179CB1	PLACFER06
47	1938744CB1	HIPONOT01
48	5557436CB1	SININOT01
49	7506178CB1	PANCTUT02
50	7506235CB1	BRSTNON02
51	1302184CB1	BRAIFER05
52	7506232CB1	ADRENOT08
53	2585358CB1	BRAITUT22
54	3961495CB1	BRAHTDR04
55	7500801CB1	CONNTUT04
56	7506414CB1	BRABDIE02
57	7506415CB1	BRABDIE02
58	72192179CB1	LVENNOT01
59	7505908CB1	LUNGFET03
60	6590147CB1	COLNNOT05
61	6828539CB1	FTUBTUE01
62	7170321CB1	ADREFEC01
63	7505918CB1	MLP000007
64	7505935CB1	TYMNOT04
65	4225965CB1	LATRTUT02
66	7495594CB1	PLACNOB01

Table 6

Library	Vector	Library Description
ADREFEC01	pINCY	This large size-fractionated library was constructed using RNA isolated from adrenal tissue removed from a Caucasian female fetus who died from anencephalus after 16-weeks' gestation. Serology was negative. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother.
ADRENOT08	pINCY	Library was constructed using RNA isolated from adrenal tissue removed from a 20-year-old Caucasian male, who died from head trauma.
BMARTXT02	pINCY	Library was constructed using RNA isolated from treated SH-SY5Y cell line derived from bone marrow neuroblastoma tumor cells removed from a 4-year-old Caucasian female. The cells were cultured in the presence of retinoic acid.
BRABDIE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male who died from a cerebrovascular accident. Serologies were negative. Patient history included Huntington's disease, emphysema, and tobacco abuse (3-4 packs per day, for 40 years).
BRAHTDR04	PCDNA2.1	This random primed library was constructed using RNA isolated from archaetocortex, anterior hippocampus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAIFER05	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAITUT22	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the right frontal/parietal lobe of a 76-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a meningioma. Family history included senile dementia.

Table 6

Library	Vector	Library Description
BRSTNON02	pINCY	This normalized breast tissue library was constructed from 6.2 million independent clones from a pool of two libraries from two different donors. Starting RNA was made from breast tissue removed from a 46-year-old Caucasian female during a bilateral reduction mammoplasty (donor A), and from breast tissue removed from a 60-year-old Caucasian female during a bilateral reduction mammoplasty (donor B). Pathology indicated normal breast parenchyma, bilaterally (A) and bilateral mammary hypertrophy (B). Patient history included hypertrophy of breast, obesity, lumbago, and glaucoma (A) and joint pain in the shoulder, thyroid cyst, colon cancer, normal delivery and cervical cancer (B). Family history included cataract, osteoarthritis, uterine cancer, benign hypertension, hyperlipidemia, and alcoholic cirrhosis of the liver, cerebrovascular disease, and type II diabetes (A) and cerebrovascular accident, atherosclerotic coronary artery disease, colon cancer, type II diabetes, hyperlipidemia, depressive disorder, and Alzheimer's Disease. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
CERVNOT01	PSPORT1	Library was constructed using RNA isolated from the uterine cervical tissue of a 35-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. Pathology indicated mild chronic cervicitis. Family history included atherosclerotic coronary artery disease and type II diabetes.
COLNNOT05	PSPORT1	Library was constructed using RNA isolated from the sigmoid colon tissue of a 40-year-old Caucasian male during a partial colectomy. Pathology indicated Crohn's disease involving the proximal colon and including the cecum. The ascending and transverse colon displayed linear ulcerations and skip lesions. There was transmural inflammation but no fistulas.
CONNTUT04	pINCY	Library was constructed using RNA isolated from tumorous spinal tissue removed from a 35-year-old Caucasian male during an exploratory laparotomy. Pathology indicated schwannoma with degenerative changes. Patient history included anxiety, depression, neurofibromatosis and benign neoplasm of the scrotum. Previously the patient had a spinal fusion. Family history included brain cancer, liver disease, and multiple sclerosis.
EOSIHET02	PBLUESCRIPT	Library was constructed using RNA isolated from peripheral blood cells apheresed from a 48-year-old Caucasian male. Patient history included hypereosinophilia. The cell population was determined to be greater than 77% eosinophils by Wright's staining.

Table 6

Library	Vector	Library Description
FIBRTXS07	pINCY	This subtracted library was constructed using 1.3 million clones from a dermal fibroblast library and was subjected to two rounds of subtraction hybridization with 2.8 million clones from an untreated dermal fibroblast tissue library. The starting library for subtraction was constructed using RNA isolated from treated dermal fibroblast tissue removed from the breast of a 31-year-old Caucasian female. The cells were treated with 9CIS retinoic acid. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated dermal fibroblast tissue from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al., Genome Research (1996) 6:791.
FTUBTUE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from right fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid (80%) and serous (20%) adenocarcinoma of the right fallopian tube, which was confined to the mucosa without mural involvement. Endometrioid carcinoma in situ was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma in situ and moderately differentiated invasive adenocarcinoma arising in an endometrial polyp. A metastatic endometrioid and serous adenocarcinoma was present in the cul-de-sac tumor. The patient presented with a pelvic mass and ascites. Patient history included medullary carcinoma of the thyroid and myocardial infarction. Patient medications included Nitro-Dur, Lescol, Lasix and Cardizem.
HEAONOR01	pINCY	This random primed library was constructed using pooled RNA isolated from aorta tissue removed from a 10-year-old Caucasian male (donor A) who died from anoxia and a 27-year-old Caucasian female (donor B) who died from an intracranial bleed. Patient history included asthma and suicidal tendency in donor A, and cerebral agenesis in donor B.
HIPONOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the hippocampus tissue of a 72-year-old Caucasian female who died from an intracranial bleed. Patient history included nose cancer, hypertension, and arthritis.
LATRUT02	pINCY	Library was constructed using RNA isolated from a myxoma removed from the left atrium of a 43-year-old Caucasian male during annuloplasty. Pathology indicated atrial myxoma. Patient history included pulmonary insufficiency, acute myocardial infarction, atherosclerotic coronary artery disease, hyperlipidemia, and tobacco use. Family history included benign hypertension, acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.

Table 6

Library	Vector	Library Description
LIVRNON08	pINCY	This normalized library was constructed from 5.7 million independent clones from a pooled liver tissue library. Starting RNA was made from pooled liver tissue removed from a 4-year-old Hispanic male who died from anoxia and a 16 week female fetus who died after 16-weeks gestation from anencephaly. Serologies were positive for cytomegalovirus in the 4-year-old. Patient history included asthma in the 4-year-old. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother of the fetus. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LUNGFEET03	pINCY	Library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
LUNGTUT17	pINCY	Library was constructed using RNA isolated from lung tumor tissue removed from a 53-year-old male. Pathology indicated grade 4 adenocarcinoma.
LVENNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the left ventricle of a 51-year-old Caucasian female, who died from an intracranial bleed.
MUSLTDR02	PCDNA2.1	This random primed library was constructed using RNA isolated from right lower thigh muscle tissue removed from a 58-year-old Caucasian male during a wide resection of the right posterior thigh. Pathology indicated no residual tumor was identified in the right posterior thigh soft tissue. Changes were consistent with a previous biopsy site. On section through the soft tissue and muscle there was a smooth cystic cavity with hemorrhage around the margin on one side. The wall of the cyst was smooth and pale-tan. Pathology for the matched tumor tissue indicated a grade II liposarcoma. Patient history included liposarcoma (right thigh), and hypercholesterolemia. Previous surgeries included resection of right thigh mass. Family history included myocardial infarction and an unspecified rare blood disease.
NOEDIT02	pINCY	Library was constructed using RNA isolated from nasal polyp tissue.
PANCTUT02	pINCY	Library was constructed using RNA isolated from pancreatic tumor tissue removed from a 45-year-old Caucasian female during radical pancreaticoduodenectomy. Pathology indicated a grade 4 anaplastic carcinoma. Family history included benign hypertension, hyperlipidemia and atherosclerotic coronary artery disease.
PLACFER01	pINCY	The library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus, who died after 16 weeks' gestation from fetal demise and hydrocephalus. Patient history included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV. Family history included multiple pregnancies and live births, and an abortion.

Table 6

Library	Vector	Library Description
PLACFER06	pINCY	This random primed library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus who died after 16 weeks' gestation from fetal demise and hydrocephalus. Patient history included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV. Family history included multiple pregnancies and live births, and an abortion.
PLACNOB01	PBLUESCRIPT	Library was constructed using RNA isolated from placenta.
SININOT01	pINCY	Library was constructed using RNA isolated from ileum tissue obtained from the small intestine of a 4-year-old Caucasian female, who died from a closed head injury. Serologies were negative. Patient history included jaundice. Previous surgeries included a double hernia repair.
TLYMNOT04	pINCY	Library was constructed using RNA isolated from activated Th1 cells. These cells were differentiated from umbilical cord CD4 T cells with IL-12 and B7-transfected COS cells, and then activated for six hours with anti-CD3 and anti-CD28 antibodies.



Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value=1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity=95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. & S. Henikoff (1996) Methods Enzymol. 266:88-105; and Altwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART, and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, p. 1-350	PFAM, INCY, SMART, or TIGRFAM hits: Probability value=1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
fileScan Pro	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
ed Phr	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phra p	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
C n o s d P S e S c an	A graphical tool for viewing and editing Phrap assemblies. A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Gordon, D. et al. (1998) Genome Res. 8:195-202. Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
MAP T	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
T	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
M t o . s if	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
47	1938744	4884960H1	SNP00003355	184	287	C	C	T	D13	n/a	n/a	n/a	n/a
48	5557436	1355739H1	SNP00007022	141	1140	A	A	G	T358	0.76	n/a	n/a	n/a
48	5557436	1458168H1	SNP00062126	149	902	G	G	T	R279	n/a	n/a	n/a	n/a
48	5557436	1458168H1	SNP00150968	157	910	A	A	G	K282	n/a	n/a	n/a	n/a
48	5557436	1559353H1	SNP00104958	137	979	C	C	A	P305	n/d	n/a	n/a	n/a
48	5557436	1877350H1	SNP00031984	79	906	G	G	A	M280	n/a	n/a	n/a	n/a
48	5557436	2658335H1	SNP00104959	169	1448	G	G	C	S461	0.95	n/a	n/a	n/a
48	5557436	4914786H1	SNP00018622	77	190	T	T	C	L42	n/a	n/a	n/a	n/a
48	5557436	518804H1	SNP00031985	146	1486	T	T	C	S474	n/a	n/a	n/a	n/a
48	5557436	518804H1	SNP00104959	102	1442	G	G	C	G459	0.95	n/a	n/a	n/a
48	5557436	6355958H1	SNP00018622	180	180	T	T	C	H38	n/a	n/a	n/a	n/a
48	5557436	6825671J1	SNP00031985	26	1468	T	T	C	S468	n/a	n/a	n/a	n/a
48	5557436	6825671J1	SNP00104959	69	1424	G	G	C	C453	0.95	n/a	n/a	n/a
48	5557436	7070058H1	SNP00018622	187	556	T	T	C	F164	n/a	n/a	n/a	n/a
48	5557436	7074007H1	SNP00149258	79	1064	C	C	T	A333	n/a	n/a	n/a	n/a
48	5557436	7470301H1	SNP00138788	414	431	T	T	C	V122	n/a	n/a	n/a	n/a
48	5557436	833625H1	SNP00042569	100	218	G	G	A	W51	n/d	n/a	n/a	n/a
49	7506178	002716H1	SNP00055501	236	1171	T	T	G	noncoding	n/a	n/a	n/a	n/a
49	7506178	1470902H1	SNP00055502	24	614	T	T	G	W115	n/a	n/a	n/a	n/a
49	7506178	2240321H1	SNP00061911	47	94	G	G	T	noncoding	n/a	n/a	n/a	n/a
49	7506178	2322443H1	SNP00137268	72	578	C	C	T	P103	n/a	n/a	n/a	n/a
49	7506178	2553222H1	SNP00140217	88	982	G	G	G	E237	n/a	n/a	n/a	n/a
49	7506178	5838342H1	SNP00014892	216	1307	G	G	A	noncoding	n/a	n/a	n/a	n/a
49	7506178	7081042H1	SNP00000200	73	1167	A	A	G	noncoding	n/a	n/a	n/a	n/a
49	7506178	7081042H1	SNP00014893	67	1173	A	A	G	noncoding	n/a	n/a	n/a	n/a
50	7506235	1892631H1	SNP00140072	64	68	T	T	G	noncoding	n/a	n/a	n/a	n/a
50	7506235	1892631H1	SNP00140073	105	109	G	G	A	noncoding	n/a	n/a	n/a	n/a
50	7506235	2963354H1	SNP00154800	70	448	A	A	G	N95	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
50	7506235	3501309H1	SNP00139135	71	1409	C	T	C	noncoding	n/a	n/a	n/a	n/a
50	7506235	5842691H1	SNP00120235	260	1067	G	A	G	noncoding	n/a	n/a	n/a	n/a
51	1302184	6987094H1	SNP00065706	295	3575	C	C	A	P1130	n/a	n/a	n/a	n/a
52	7506232	1281980H1	SNP00052301	225	193	T	T	C	S28	n/d	n/d	n/d	n/d
53	2585358	1238586H1	SNP00044373	85	668	C	C	G	L97	n/a	n/a	n/a	n/a
53	2585358	1238586H1	SNP00140452	60	643	A	A	C	H89	n/a	n/a	n/a	n/a
53	2585358	1532857H1	SNP00018146	154	828	C	C	G	noncoding	n/a	n/a	n/a	n/a
54	3961495	1568780H1	SNP00044374	91	887	G	G	T	noncoding	n/a	n/a	n/a	n/a
54	3961495	6076413H1	SNP00104670	93	2236	C	C	A	P718	n/a	n/a	n/a	n/a
55	7500801	2705502H1	SNP000035039	139	4445	A	A	G	S1455	n/d	n/d	n/d	n/a
55	7500801	3385851H1	SNP00000177	291	987	T	C	T	H152	0.93	n/a	n/a	n/a
55	7500801	6330582H1	SNP00058790	122	1244	C	C	T	A238	n/d	n/d	0.99	n/d
55	7500801	6753392H1	SNP00031881	303	2208	C	C	A	noncoding	n/a	n/a	n/a	n/a
58	72192179	186759H1	SNP00058791	287	1365	C	G	C	P278	n/a	n/a	n/a	n/a
58	72192179	2635387H1	SNP00016115	124	3068	A	A	G	I543	0.90	0.76	0.91	0.88
58	72192179	2635387H1	SNP00128590	124	5512	T	T	G	noncoding	n/a	n/a	n/a	n/a
58	72192179	3402795H1	SNP00128591	183	5571	G	G	C	noncoding	0.64	n/a	n/a	n/a
58	72192179	5278841H1	SNP00016115	196	3066	G	A	G	G542	0.90	0.76	0.91	0.88
58	72192179	5278841H1	SNP00128590	140	5500	T	T	G	noncoding	n/a	n/a	n/a	n/a
58	72192179	5492831H1	SNP00128590	199	5559	G	G	C	noncoding	0.64	n/a	n/a	n/a
58	72192179	5492831H1	SNP00128591	87	5504	T	T	G	noncoding	n/a	n/a	n/a	n/a
58	72192179	6906413H1	SNP00072143	385	164	A	G	C	noncoding	0.64	n/a	n/a	n/a
58	72192179	8125668H1	SNP00127050	285	1725	C	C	A	noncoding	n/d	n/a	n/a	n/a
60	6590147	1307146H1	SNP00141393	85	58	C	C	T	P95	n/a	n/a	n/a	n/a
60	6590147	1359153H1	SNP00110640	225	1305	A	A	C	noncoding	n/a	n/a	n/a	n/a
60	6590147	1669205H1	SNP00065202	152	393	A	A	G	V70	n/a	n/a	n/a	n/a
60	6590147	1711357H1	SNP00110640	67	1304	A	A	C	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
60	6590147	1746237H1	SNP00110640	78	1303	A	A	C	noncoding	n/a	n/a	n/a	n/a
60	6590147	2298510H1	SNP00110640	72	1299	A	A	C	noncoding	n/a	n/a	n/a	n/a
60	6590147	2442071H1	SNP00110640	165	1306	A	A	C	noncoding	n/a	n/a	n/a	n/a
60	6590147	2968840H1	SNP00065202	12	391	A	A	G	I70	n/a	n/a	n/a	n/a
60	6590147	3210518H1	SNP00065202	11	390	G	A	G	K69	n/a	n/a	n/a	n/a
60	6590147	3225413H1	SNP00110640	201	1301	A	A	C	noncoding	n/a	n/a	n/a	n/a
60	6590147	3440973H1	SNP00110640	86	1287	A	A	C	noncoding	n/a	n/a	n/a	n/a
60	6590147	3484802H1	SNP00110640	151	1294	A	A	C	noncoding	n/a	n/a	n/a	n/a
60	6590147	3527267H1	SNP00065202	17	392	A	A	G	E70	n/a	n/a	n/a	n/a
60	6590147	3537755H1	SNP00110640	80	1302	A	A	C	noncoding	n/a	n/a	n/a	n/a
60	6590147	4211251H1	SNP00141393	17	57	C	C	A	noncoding	n/a	n/a	n/a	n/a
60	6590147	6755049H1	SNP00110640	404	1312	A	A	C	noncoding	n/a	n/a	n/a	n/a
61	6828539	546676H1	SNP00015375	225	1516	G	C	G	R489	n/a	n/a	n/a	n/a
61	6828539	6082669H1	SNP00015375	222	1515	C	C	G	Q489	n/a	n/a	n/a	n/a
62	7170321	1361284H1	SNP00127502	125	223	G	G	A	R59	n/a	n/a	n/a	n/a
62	7170321	1390974H1	SNP00127502	236	224	G	G	A	R59	n/a	n/a	n/a	n/a
62	7170321	1606071H1	SNP00009850	57	1462	G	G	A	noncoding	0.99	n/d	n/d	n/a
62	7170321	1889073H1	SNP00009850	121	1461	G	G	A	noncoding	0.99	n/d	n/d	n/a
62	7170321	2499608H1	SNP00127502	34	222	G	G	A	G59	n/a	n/a	n/a	n/a
62	7170321	2670638H1	SNP00009850	149	1465	G	G	A	noncoding	0.99	n/d	n/d	n/a
62	7170321	2676623H1	SNP00009850	90	1458	G	G	A	noncoding	0.99	n/d	n/d	n/a
62	7170321	2959873H1	SNP00127502	219	218	G	G	A	W57	n/a	n/a	n/a	n/a
62	7170321	3093844H1	SNP00127502	196	220	G	G	A	R58	n/a	n/a	n/a	n/a
62	7170321	3163826H1	SNP00127502	216	221	G	G	A	K58	n/a	n/a	n/a	n/a
62	7170321	3277346H1	SNP00127502	219	219	G	G	A	E58	n/a	n/a	n/a	n/a
62	7170321	3350415H1	SNP00127502	193	212	G	G	A	R55	n/a	n/a	n/a	n/a
62	7170321	3492203H1	SNP00009850	203	1460	G	G	A	noncoding	0.99	n/d	n/d	n/a
62	7170321	3644419H1	SNP00009850	165	1459	G	G	A	noncoding	0.99	n/d	n/d	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
62	7170321	3804929H1	SNP00127502	192	207	G	G	A	G54	n/a	n/a	n/a	n/a
62	7170321	3814747H1	SNP00127502	195	216	G	G	A	G57	n/a	n/a	n/a	n/a
62	7170321	3967036H1	SNP00009850	73	1467	G	G	A	noncoding	0.99	n/d	n/d	n/a
62	7170321	4891916H1	SNP00127502	187	176	G	G	A	L43	n/a	n/a	n/a	n/a
62	7170321	5104105H1	SNP00009850	131	1463	G	G	A	noncoding	0.99	n/d	n/d	n/a
62	7170321	5197323H1	SNP00009850	127	1453	G	G	A	noncoding	0.99	n/d	n/d	n/a
62	7170321	5767980H1	SNP00127502	208	374	G	G	A	E109	n/a	n/a	n/a	n/a
62	7170321	5956262H1	SNP00127502	220	379	G	G	A	S111	n/a	n/a	n/a	n/a
62	7170321	6151461H1	SNP00009850	185	1450	G	G	A	noncoding	0.99	n/d	n/d	n/a
62	7170321	6217828H1	SNP00127502	189	368	G	G	A	K107	n/a	n/a	n/a	n/a
62	7170321	6403414H1	SNP00127502	220	229	G	G	A	G61	n/a	n/a	n/a	n/a
62	7170321	640639H1	SNP00127502	212	217	G	G	A	C57	n/a	n/a	n/a	n/a
62	7170321	6816552J1	SNP00127502	267	373	G	G	A	G109	n/a	n/a	n/a	n/a
63	7505918	1877128H1	SNP00107685	91	2017	A	A	G	M642	n/d	n/d	n/d	n/d
63	7505918	2266413H1	SNP00126002	215	647	C	C	T	S185	n/a	n/a	n/a	n/a
63	7505918	2316120H1	SNP00049166	144	2445	C	C	G	noncoding	0.98	n/a	n/a	n/a
63	7505918	4227736H1	SNP00049166	36	2440	C	C	G	noncoding	0.98	n/a	n/a	n/a
63	7505918	4638868H1	SNP00107685	228	2016	A	A	G	L641	n/d	n/d	n/d	n/d
63	7505918	4691506H1	SNP00126002	181	648	C	C	T	Y185	n/a	n/a	n/a	n/a
63	7505918	4772413H1	SNP00049166	77	2444	C	C	G	noncoding	0.98	n/a	n/a	n/a
63	7505918	5643220H1	SNP00107684	170	476	C	C	T	S128	n/a	n/a	n/a	n/a
63	7505918	5972195H1	SNP00033129	159	2349	T	T	G	F752	n/d	n/d	n/d	n/d
63	7505918	6772406J1	SNP00107684	442	477	T	T	C	Y128	n/a	n/a	n/a	n/a
63	7505918	6890125J1	SNP00033128	146	2292	T	T	C	A733	n/a	n/a	n/a	n/a
63	7505918	7675370H1	SNP00126002	314	650	T	C	T	V186	n/a	n/a	n/a	n/a
64	7505935	2100266H1	SNP00052637	44	2151	C	C	T	L676	n/d	n/a	n/a	n/a
64	7505935	2791525H1	SNP00095095	170	2775	T	C	T	L884	n/d	n/d	n/d	n/d
64	7505935	2894643H1	SNP00052637	52	2146	C	C	T	P674	n/d	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
64	7505935	5972523H1	SNP00095095	119	2779	T	C	T	I885	n/d	n/d	n/d	n/d
64	7505935	6819222J1	SNP00114608	475	301	C	C	T	P59	n/d	n/d	n/d	n/d
64	7505935	6827687J1	SNP00052637	18	2148	C	C	T	L675	n/d	n/a	n/a	n/a
65	4225965	1222633H1	SNP00028248	229	2353	C	C	T	C656	n/a	n/a	n/a	n/a
65	4225965	3340114H1	SNP00146019	199	763	C	C	T	I126	n/a	n/a	n/a	n/a
65	4225965	3579326H1	SNP00146018	236	285	A	G	A	noncoding	n/a	n/a	n/a	n/a
65	4225965	5960727H1	SNP00146019	310	759	C	C	T	S125	n/a	n/a	n/a	n/a
65	4225965	6826774H1	SNP00140458	89	427	C	C	T	S14	n/a	n/a	n/a	n/a
65	4225965	6826774J1	SNP00140458	25	426	C	C	T	T14	n/a	n/a	n/a	n/a
65	4225965	7152528H1	SNP00146019	195	762	C	C	T	T126	n/a	n/a	n/a	n/a
65	4225965	852987H1	SNP00028248	92	2360	C	C	T	L659	n/a	n/a	n/a	n/a
66	7495594	062963H1	SNP00033362	251	10525	G	G	A	G3410	n/a	n/a	n/a	n/a
66	7495594	1308750H1	SNP00033363	171	10848	C	C	T	Y3517	n/a	n/a	n/a	n/a
66	7495594	1516984H1	SNP00033362	64	10526	G	G	A	C3410	n/a	n/a	n/a	n/a
66	7495594	2210156H1	SNP00053764	137	10012	T	T	G	S3239	n/a	n/a	n/a	n/a
66	7495594	2210366H1	SNP00053764	136	10011	T	T	G	V3238	n/a	n/a	n/a	n/a
66	7495594	2989217H1	SNP00038892	8	2206	A	A	G	I637	n/a	n/a	n/a	n/a
66	7495594	4184585H1	SNP00033361	61	10284	C	T	C	S3329	n/d	n/d	n/d	n/d
66	7495594	4184585H1	SNP00154155	175	10398	G	C	G	L3367	n/a	n/a	n/a	n/a
66	7495594	4381289H1	SNP00033363	125	10847	C	C	T	S3517	n/a	n/a	n/a	n/a
66	7495594	4381338H1	SNP00033363	124	10846	C	C	T	H3517	n/a	n/a	n/a	n/a
66	7495594	5048485H1	SNP00033360	146	9741	T	C	T	L3148	0.52	0.63	0.83	0.49
66	7495594	5156725H1	SNP00033362	162	10527	G	G	A	W3410	n/a	n/a	n/a	n/a
66	7495594	6191181H1	SNP00033363	17	10838	C	C	T	A3514	n/a	n/a	n/a	n/a
66	7495594	6209608H1	SNP00033362	192	10516	G	G	A	G3407	n/a	n/a	n/a	n/a
66	7495594	6386095H1	SNP00033361	75	10635	T	T	C	I3446	n/d	n/d	n/d	n/d
66	7495594	6810339J1	SNP00033363	5	10841	C	C	T	A3515	n/a	n/a	n/a	n/a
66	7495594	6916010H1	SNP00071215	71	5725	C	A	C	P1810	0.83	0.98	0.63	0.74

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
66	7495594	6916051H1	SNP00071215	71	5726	C	A	C	A1810	0.83	0.98	0.63	
66	7495594	7039764H1	SNP00064305	407	1292	C	G	C	A332	0.80	0.97	0.89	0.74
66	7495594	966619H1	SNP00033362	4	10501	G	G	A	A3402	n/a	n/a	n/a	0.71
													n/a



What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6-7, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:25, SEQ ID NO:28-29, and SEQ ID NO:32-33,
  - c) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:4 and SEQ ID NO:21,
  - d) a polypeptide comprising a naturally occurring amino acid sequence at least 97% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:18 and SEQ ID NO:23-24,
  - e) a polypeptide consisting essentially of a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:9, SEQ ID NO:13-14, SEQ ID NO:19, SEQ ID NO:26-27 and SEQ ID NO:30-31,
  - f) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, and
  - g) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-33.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

- a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
- b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-33.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-66,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:34-43, SEQ ID NO:45-50, SEQ ID NO:52, SEQ ID NO:55, SEQ ID NO:58-60, SEQ ID NO:62, and SEQ ID NO:65-66,
- c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 95% identical to the polynucleotide sequence of SEQ ID NO:51,
- d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 98% identical to the polynucleotide sequence of SEQ ID NO:53,
- e) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 96% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:54 and SEQ ID NO:56-57,
- f) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 93% identical to the polynucleotide sequence of SEQ ID NO:61,

- g) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 99% identical to the polynucleotide sequence of SEQ ID NO:63,
- h) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence consisting of SEQ ID NO:64,
- i) a polynucleotide complementary to a polynucleotide of a),
- j) a polynucleotide complementary to a polynucleotide of b),
- k) a polynucleotide complementary to a polynucleotide of c),
- l) a polynucleotide complementary to a polynucleotide of d),
- m) a polynucleotide complementary to a polynucleotide of e),
- n) a polynucleotide complementary to a polynucleotide of f),
- o) a polynucleotide complementary to a polynucleotide of g),
- p) a polynucleotide complementary to a polynucleotide of h), and
- q) an RNA equivalent of a)-p).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
  - b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment

thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

5

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-33.

19. A method for treating a disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment the composition of claim 17.

10

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

15

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

20

22. A method for treating a disease or condition associated with decreased expression of functional CGDD, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

25

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

30

25. A method for treating a disease or condition associated with overexpression of functional CGDD, comprising administering to a patient in need of such treatment a composition of claim 24.

35

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- 5 b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- 10 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- 15 c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

20 28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- 25 b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- 30 a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a
- 35 polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A method for a diagnostic test for a condition or disease associated with the expression of CGDD in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of CGDD in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, further comprising a label.

35. A method of diagnosing a condition or disease associated with the expression of CGDD in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- 5
- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
  - b) isolating antibodies from the animal, and
  - c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33.

10 37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- 15
- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-33, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
  - b) isolating antibody producing cells from the animal,
  - c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
  - d) culturing the hybridoma cells, and
  - e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33.
- 20

25

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

30 42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

35

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and
- 5 b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and
- 10 b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-33.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

20 47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a
- 25 hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first

30 oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is

35 completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.



50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is  
5 completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a  
10 nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

15 55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

20

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

25 58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

30 60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

35 63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.
65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.
- 5 66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.
67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
- 10 69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
- 15 71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
- 20 74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.
75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.
- 25 76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.
77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.
78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.
- 30 79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.
80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.
- 35 81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

82. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.
83. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:28.
- 5 84. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:29.
85. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:30.
86. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:31.
- 10 87. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:32.
88. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:33.
- 15 89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.
90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.
- 20 91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.
92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.
- 25 93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.
94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.
- 30 95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.
- 35

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:45.

101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:46.

102. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:47.

103. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:48.

104. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:49.

105. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:50.

106. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:51.

107. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:52.

108. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:53.

109. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:54.

110. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:55.

111. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:56.

112. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:57.

113. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:58.

114. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:59.

115. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:60.

116. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:61.

117. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:62.

118. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:63.

119. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:64.

120. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:65.

121. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
5 NO:66.

<110> INCYTE GENOMICS, INC.  
GRIFFIN, Jennifer A  
RAMKUMAR, Jayalaxmi  
EMERLING, Brooke M.  
KABLE, Amy E.  
ELLIOTT, Vicki S.  
MARQUIS, Joseph P.  
BAUGHN, Mariah R.  
GORVAD, Ann E.  
YUE, Henry  
LEE, Ernestine A.  
BECHA, Shanya D.  
TANG, Y. Tom  
TRAN, Uyen K.  
SWARNAKAR, Anita  
LEE, Sally  
ISON, Craig H.  
HAFALIA, April J.A.  
TRAN, Bao  
SPRAGUE, William W.  
LEE, Soo Yeun  
KHARE, Reena  
GANDHI, Ameena R.  
GIETZEN, Kimberly J.  
BHATIA, Umesh  
BURRILL, John D.  
BLAKE, Julie J.  
HO, Anne  
Zheng, Wenjin

<120> PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH

<130> PF-1319 PCT

<140> To Be Assigned  
<141> Herewith

<150> US 60/340,747  
<151> 2001-12-07

<150> US 60/342,761  
<151> 2001-12-20

<150> US 60/349,705  
<151> 2002-01-15

<150> US 60/354,764  
<151> 2002-02-06

<150> US 60/356,216  
<151> 2002-02-12

<160> 66  
<170> PERL Program

<210> 1  
<211> 1104  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7722608CD1

<400> 1  
 Met Gly Asn Ala Ala Gly Ser Ala Glu Gln Pro Ala Gly Pro Ala  
 1 5 10 15  
 Ala Pro Pro Pro Lys Gln Pro Ala Pro Pro Lys Gln Pro Met Pro  
 20 25 30  
 Ala Ala Gly Glu Leu Glu Ala Arg Phe Asn Arg Ala Leu Asn Cys  
 35 40 45  
 Met Asn Leu Pro Pro Asp Lys Val Gln Leu Leu Ser Gln Tyr Asp  
 50 55 60  
 Asn Glu Lys Lys Trp Glu Leu Ile Cys Asp Gln Glu Arg Phe Gln  
 65 70 75  
 Val Lys Asn Pro Pro Ala Ala Tyr Ile Gln Lys Leu Lys Ser Tyr  
 80 85 90  
 Val Asp Thr Gly Gly Val Ser Arg Lys Val Ala Ala Asp Trp Met  
 95 100 105  
 Ser Asn Leu Gly Phe Lys Arg Arg Val Gln Glu Ser Thr Gln Val  
 110 115 120  
 Leu Arg Glu Leu Glu Thr Ser Leu Arg Thr Asn His Ile Gly Trp  
 125 130 135  
 Val Gln Glu Phe Leu Asn Glu Glu Asn Arg Gly Leu Asp Val Leu  
 140 145 150  
 Leu Glu Tyr Leu Ala Phe Ala Gln Cys Ser Val Thr Tyr Asp Met  
 155 160 165  
 Glu Ser Thr Asp Asn Gly Ala Ser Asn Ser Glu Lys Asn Lys Pro  
 170 175 180  
 Leu Glu Gln Ser Val Glu Asp Leu Ser Lys Gly Pro Pro Ser Ser  
 185 190 195  
 Val Pro Lys Ser Arg His Leu Thr Ile Lys Leu Thr Pro Ala His  
 200 205 210  
 Ser Arg Lys Ala Leu Arg Asn Ser Arg Ile Val Ser Gln Lys Asp  
 215 220 225  
 Asp Val His Val Cys Ile Met Cys Leu Arg Ala Ile Met Asn Tyr  
 230 235 240  
 Gln Ser Gly Phe Ser Leu Val Met Asn His Pro Ala Cys Val Asn  
 245 250 255  
 Glu Ile Ala Leu Ser Leu Asn Asn Lys Asn Pro Arg Thr Lys Ala  
 260 265 270  
 Leu Val Leu Glu Leu Leu Ala Ala Val Cys Leu Val Arg Gly Gly  
 275 280 285  
 His Asp Ile Ile Leu Ala Ala Phe Asp Asn Phe Lys Glu Val Cys  
 290 295 300  
 Gly Glu Gln His Arg Phe Glu Lys Leu Met Glu Tyr Phe Arg Asn  
 305 310 315  
 Glu Asp Ser Asn Ile Asp Phe Met Val Ala Cys Met Gln Phe Ile  
 320 325 330  
 Asn Ile Val Val His Ser Val Glu Asn Met Asn Phe Arg Val Phe  
 335 340 345  
 Leu Gln Tyr Glu Phe Thr His Leu Gly Leu Asp Leu Tyr Leu Glu  
 350 355 360  
 Arg Leu Arg Leu Thr Glu Ser Asp Lys Leu Gln Val Gln Ile Gln  
 365 370 375  
 Ala Tyr Leu Asp Asn Ile Phe Asp Val Gly Ala Leu Leu Glu Asp  
 380 385 390  
 Thr Glu Thr Lys Asn Ala Val Leu Glu His Met Glu Glu Leu Gln  
 395 400 405  
 Glu Gln Val Ala Leu Leu Thr Glu Arg Leu Arg Asp Ala Glu Asn  
 410 415 420  
 Glu Ser Met Ala Lys Ile Ala Glu Leu Glu Lys Gln Leu Ser Gln  
 425 430 435  
 Ala Arg Lys Glu Leu Glu Thr Leu Arg Glu Arg Phe Ser Glu Ser  
 440 445 450  
 Thr Ala Met Gly Ala Ser Arg Arg Pro Pro Glu Pro Glu Lys Ala



	455	460	465
Pro Pro Ala Ala	Pro Thr Arg Pro Ser	Ala Leu Glu Leu Lys	Val
	470	475	480
Glu Glu Leu Glu	Glu Lys Gly Leu Ile	Arg Ile Leu Arg Gly	Pro
	485	490	495
Gly Asp Ala Val	Ser Ile Glu Ile Leu	Pro Val Ala Val Ala	Thr
	500	505	510
Pro Ser Gly Gly	Asp Ala Pro Thr Pro	Gly Val Pro Thr Gly	Ser
	515	520	525
Pro Ser Pro Asp	Leu Ala Pro Ala Ala	Glu Pro Ala Pro Gly	Ala
	530	535	540
Ala Pro Pro Pro	Pro Pro Pro Leu Pro	Gly Leu Pro Ser Pro	Gln
	545	550	555
Glu Ala Pro Pro	Ser Ala Pro Pro Gln	Ala Pro Pro Leu Pro	Gly
	560	565	570
Ser Pro Glu Pro	Pro Pro Ala Pro Pro	Leu Pro Gly Asp Leu	Pro
	575	580	585
Pro Pro Pro Pro	Pro Pro Pro Pro Pro	Pro Gly Thr Asp Gly	Pro
	590	595	600
Val Pro Pro Pro	Pro Pro Pro Pro Pro	Pro Pro Pro Gly Gly	Pro
	605	610	615
Pro Asp Ala Leu	Gly Arg Arg Asp Ser	Glu Leu Gly Pro Gly	Val
	620	625	630
Lys Ala Lys Lys	Pro Ile Gln Thr Lys	Phe Arg Met Pro Leu	Leu
	635	640	645
Asn Trp Val Ala	Leu Lys Pro Ser Gln	Ile Thr Gly Thr Val	Phe
	650	655	660
Thr Glu Leu Asn	Asp Glu Lys Val Leu	Gln Glu Leu Asp Met	Ser
	665	670	675
Asp Phe Glu Glu	Gln Phe Lys Thr Lys	Ser Gln Gly Pro Ser	Leu
	680	685	690
Asp Leu Ser Ala	Leu Lys Ser Lys Ala	Ala Gln Lys Ala Pro	Ser
	695	700	705
Lys Ala Thr Leu	Ile Glu Ala Asn Arg	Ala Lys Asn Leu Ala	Ile
	710	715	720
Thr Leu Arg Lys	Gly Asn Leu Gly Ala	Glu Arg Ile Cys Gln	Ala
	725	730	735
Ile Glu Ala Tyr	Asp Leu Gln Ala Leu	Gly Leu Asp Phe Leu	Glu
	740	745	750
Leu Leu Met Arg	Phe Leu Pro Thr Glu	Tyr Glu Arg Ser Leu	Ile
	755	760	765
Thr Arg Phe Glu	Arg Glu Gln Arg Pro	Met Glu Glu Leu Ser	Glu
	770	775	780
Glu Asp Arg Phe	Met Leu Cys Phe Ser	Arg Ile Pro Arg Leu	Pro
	785	790	795
Glu Arg Met Thr	Thr Leu Thr Phe Leu	Gly Asn Phe Pro Asp	Thr
	800	805	810
Ala Gln Leu Leu	Met Pro Gln Leu Asn	Ala Ile Ile Ala Ala	Ser
	815	820	825
Met Ser Ile Lys	Ser Ser Asp Lys Leu	Arg Gln Ile Leu Glu	Ile
	830	835	840
Val Leu Ala Phe	Gly Asn Tyr Met Asn	Ser Ser Lys Arg Gly	Ala
	845	850	855
Ala Tyr Gly Phe	Arg Leu Gln Ser Leu	Asp Ala Leu Leu Glu	Met
	860	865	870
Lys Ser Thr Asp	Arg Lys Gln Thr Leu	Leu His Tyr Leu Val	Lys
	875	880	885
Val Ile Ala Glu	Lys Tyr Pro Gln Leu	Thr Gly Phe His Ser	Asp
	890	895	900
Leu His Phe Leu	Asp Lys Ala Gly Ser	Val Ser Leu Asp Ser	Val
	905	910	915
Leu Ala Asp Val	Arg Ser Leu Gln Arg	Gly Leu Glu Leu Thr	Gln
	920	925	930

Arg Glu Phe Val Arg Gln Asp Asp Cys Met Val Leu Lys Glu Phe  
 935 940 945  
 Leu Arg Ala Asn Ser Pro Thr Met Asp Lys Leu Leu Ala Asp Ser  
 950 955 960  
 Lys Thr Ala Gln Glu Ala Phe Glu Ser Val Val Glu Tyr Phe Gly  
 965 970 975  
 Glu Asn Pro Lys Thr Thr Ser Pro Gly Leu Phe Phe Ser Leu Phe  
 980 985 990  
 Ser Arg Phe Ile Lys Ala Tyr Lys Lys Ala Glu Gln Glu Val Glu  
 995 1000 1005  
 Gln Trp Lys Lys Glu Ala Ala Ala Gln Glu Ala Gly Ala Asp Thr  
 1010 1015 1020  
 Pro Gly Lys Gly Glu Pro Pro Ala Pro Lys Ser Pro Pro Lys Ala  
 1025 1030 1035  
 Arg Arg Pro Gln Met Asp Leu Ile Ser Glu Leu Lys Arg Arg Gln  
 1040 1045 1050  
 Gln Lys Glu Pro Leu Ile Tyr Glu Ser Asp Arg Asp Gly Ala Ile  
 1055 1060 1065  
 Glu Asp Ile Ile Thr Asp Leu Arg Asn Gln Pro Tyr Ile Arg Ala  
 1070 1075 1080  
 Asp Thr Gly Arg Arg Ser Ala Arg Arg Arg Pro Pro Gly Pro Pro  
 1085 1090 1095  
 Leu Gln Val Thr Ser Asp Leu Ser Leu  
 1100

<210> 2  
 <211> 1299  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7505869CD1

<400> 2  
 Met Lys Val Phe Ser Leu Ile Gln Gln Ala Val Thr Arg Tyr Arg  
 1 5 10 15  
 Lys Ala Ile Ala Lys Asp Pro Asn Tyr Trp Ile Gln Val His Arg  
 20 25 30  
 Leu Glu His Gly Asp Gly Gly Ile Leu Asp Leu Asp Asp Ile Leu  
 35 40 45  
 Cys Asp Val Ala Asp Asp Lys Asp Arg Leu Val Ala Val Phe Asp  
 50 55 60  
 Glu Gln Asp Pro His His Gly Gly Asp Gly Thr Ser Ala Ser Ser  
 65 70 75  
 Thr Gly Thr Gln Ser Pro Glu Ile Phe Gly Ser Glu Leu Gly Thr  
 80 85 90  
 Asn Asn Val Ser Ala Phe Gln Pro Tyr Gln Ala Thr Ser Glu Ile  
 95 100 105  
 Glu Val Thr Pro Ser Val Leu Arg Ala Asn Met Pro Leu His Val  
 110 115 120  
 Arg Arg Ser Ser Asp Pro Ala Leu Ile Gly Leu Ser Thr Ser Val  
 125 130 135  
 Ser Asp Ser Asn Phe Ser Ser Glu Glu Pro Ser Arg Lys Asn Pro  
 140 145 150  
 Thr Arg Trp Ser Thr Thr Ala Gly Phe Leu Lys Gln Asn Thr Ala  
 155 160 165  
 Gly Ser Pro Lys Thr Cys Asp Arg Lys Lys Asp Glu Asn Tyr Arg  
 170 175 180  
 Ser Leu Pro Arg Asp Thr Ser Asn Trp Ser Asn Gln Phe Gln Arg  
 185 190 195  
 Asp Asn Ala Arg Ser Ser Leu Ser Ala Ser His Pro Met Val Gly  
 200 205 210

Lys Trp Leu Glu Lys Gln Glu Gln Asp Glu Asp Gly Thr Glu Glu  
 215 220 225  
 Asp Asn Ser Arg Val Glu Pro Val Gly His Ala Asp Thr Gly Leu  
 230 235 240  
 Glu His Ile Pro Asn Phe Ser Leu Asp Asp Met Val Lys Leu Val  
 245 250 255  
 Glu Val Pro Asn Asp Gly Gly Pro Leu Gly Ile His Val Val Pro  
 260 265 270  
 Phe Ser Ala Arg Gly Gly Arg Thr Leu Gly Leu Leu Val Lys Arg  
 275 280 285  
 Leu Glu Lys Gly Gly Lys Ala Glu His Glu Asn Leu Phe Arg Glu  
 290 295 300  
 Asn Asp Cys Ile Val Arg Ile Asn Asp Gly Asp Leu Arg Asn Arg  
 305 310 315  
 Arg Phe Glu Gln Ala Gln His Met Phe Arg Gln Ala Met Arg Thr  
 320 325 330  
 Pro Ile Ile Trp Phe His Val Val Pro Ala Ala Asn Lys Glu Gln  
 335 340 345  
 Tyr Glu Gln Leu Ser Gln Ser Glu Lys Asn Asn Tyr Tyr Ser Ser  
 350 355 360  
 Arg Phe Ser Pro Asp Ser Gln Tyr Ile Asp Asn Arg Ser Val Asn  
 365 370 375  
 Ser Ala Gly Leu His Thr Val Gln Arg Ala Pro Arg Leu Asn His  
 380 385 390  
 Pro Pro Glu Gln Ile Asp Ser His Ser Arg Leu Pro His Ser Ala  
 395 400 405  
 His Pro Ser Gly Lys Pro Pro Ser Ala Pro Ala Ser Ala Pro Gln  
 410 415 420  
 Asn Val Phe Ser Thr Thr Val Ser Ser Gly Tyr Asn Thr Lys Lys  
 425 430 435  
 Ile Gly Lys Arg Leu Asn Ile Gln Leu Lys Lys Gly Thr Glu Gly  
 440 445 450  
 Leu Gly Phe Ser Ile Thr Ser Arg Asp Val Thr Ile Gly Gly Ser  
 455 460 465  
 Ala Pro Ile Tyr Val Lys Asn Ile Leu Pro Arg Gly Ala Ala Ile  
 470 475 480  
 Gln Asp Gly Arg Leu Lys Ala Gly Asp Arg Leu Ile Glu Val Asn  
 485 490 495  
 Gly Val Asp Leu Val Gly Lys Ser Gln Glu Glu Val Val Ser Leu  
 500 505 510  
 Leu Arg Ser Thr Lys Met Glu Gly Thr Val Ser Leu Leu Val Phe  
 515 520 525  
 Arg Gln Glu Asp Ala Phe His Pro Arg Glu Leu Asn Ala Glu Pro  
 530 535 540  
 Ser Gln Met Gln Ile Pro Lys Glu Thr Lys Ala Glu Asp Glu Asp  
 545 550 555  
 Ile Val Leu Thr Pro Asp Gly Thr Arg Glu Phe Leu Thr Phe Glu  
 560 565 570  
 Val Pro Leu Asn Asp Ser Gly Ser Ala Gly Leu Gly Val Ser Val  
 575 580 585  
 Lys Gly Asn Arg Ser Lys Glu Asn His Ala Asp Leu Gly Ile Phe  
 590 595 600  
 Val Lys Ser Ile Ile Asn Gly Gly Ala Ala Ser Lys Asp Gly Arg  
 605 610 615  
 Leu Arg Val Asn Asp Gln Leu Ile Ala Val Asn Gly Glu Ser Leu  
 620 625 630  
 Leu Gly Lys Thr Asn Gln Asp Ala Met Glu Thr Leu Arg Arg Ser  
 635 640 645  
 Met Ser Thr Glu Gly Asn Lys Arg Gly Met Ile Gln Leu Ile Val  
 650 655 660  
 Ala Arg Arg Ile Ser Lys Cys Asn Glu Leu Lys Ser Pro Gly Ser  
 665 670 675  
 Pro Pro Gly Pro Glu Leu Pro Ile Glu Thr Ala Leu Asp Asp Arg

	680		685		690
Glu Arg Arg Ile	Ser His Ser Leu Tyr	Ser Gly Ile Glu Gly	Leu		
	695		700		705
Asp Glu Ser Pro	Ser Arg Asn Ala Ala	Leu Ser Arg Ile Met	Gly		
	710		715		720
Glu Ser Gly Lys	Tyr Gln Leu Ser Pro	Thr Val Asn Met Pro	Gln		
	725		730		735
Asp Asp Thr Val	Ile Ile Glu Asp Asp	Arg Leu Pro Val Leu	Pro		
	740		745		750
Pro His Leu Ser	Asp Gln Ser Ser Ser	Ser Ser His Asp Asp	Val		
	755		760		765
Gly Phe Val Thr	Ala Asp Ala Gly Thr	Trp Ala Lys Ala Ala	Ile		
	770		775		780
Ser Asp Ser Ala	Asp Cys Ser Leu Ser	Pro Asp Val Asp Pro	Val		
	785		790		795
Leu Ala Phe Gln	Arg Glu Gly Phe Gly	Arg Gln Ser Met Ser	Glu		
	800		805		810
Lys Arg Thr Lys	Gln Phe Ser Asp Ala	Ser Gln Leu Asp Phe	Val		
	815		820		825
Lys Thr Arg Lys	Ser Lys Ser Met Asp	Leu Gly Ile Ala Asp	Glu		
	830		835		840
Thr Lys Leu Asn	Thr Val Asp Asp Gln	Lys Ala Gly Ser Pro	Ser		
	845		850		855
Arg Asp Val Gly	Pro Ser Leu Gly Leu	Lys Lys Ser Ser Ser	Leu		
	860		865		870
Glu Ser Leu Gln	Thr Ala Val Ala Glu	Val Thr Leu Asn Gly	Asp		
	875		880		885
Ile Pro Phe His	Arg Pro Arg Pro Arg	Ile Ile Arg Gly Arg	Gly		
	890		895		900
Cys Asn Glu Ser	Phe Arg Ala Ala Ile	Asp Lys Ser Tyr Asp	Lys		
	905		910		915
Pro Ala Val Asp	Asp Asp Asp Glu Gly	Met Glu Thr Leu Glu	Glu		
	920		925		930
Asp Thr Glu Glu	Ser Ser Arg Ser Gly	Arg Glu Ser Val Ser	Thr		
	935		940		945
Ala Ser Asp Gln	Pro Ser His Ser Leu	Glu Arg Gln Met Asn	Gly		
	950		955		960
Asn Gln Glu Lys	Gly Asp Lys Thr Asp	Arg Lys Lys Asp Lys	Thr		
	965		970		975
Gly Lys Glu Lys	Lys Lys Asp Arg Asp	Lys Glu Lys Asp Lys	Met		
	980		985		990
Lys Ala Lys Lys	Gly Met Leu Lys Gly	Leu Gly Asp Met Phe	Arg		
	995		1000		1005
Ile Gln Ala Lys	Thr Arg Glu Phe Arg	Glu Arg Gln Ala Arg	Glu		
	1010		1015		1020
Arg Asp Tyr Ala	Glu Ile Gln Asp Phe	His Arg Thr Phe Gly	Cys		
	1025		1030		1035
Asp Asp Glu Leu	Met Tyr Gly Gly Val	Ser Ser Tyr Glu Gly	Ser		
	1040		1045		1050
Met Ala Leu Asn	Ala Arg Pro Gln Ser	Pro Arg Glu Gly His	Met		
	1055		1060		1065
Met Asp Ala Leu	Tyr Ala Gln Val Lys	Lys Pro Arg Asn Ser	Lys		
	1070		1075		1080
Pro Ser Pro Val	Asp Ser Asn Arg Ser	Thr Pro Ser Asn His	Asp		
	1085		1090		1095
Arg Ile Gln Arg	Leu Arg Gln Glu Phe	Gln Gln Ala Lys Gln	Asp		
	1100		1105		1110
Glu Asp Val Glu	Asp Arg Arg Arg Thr	Tyr Ser Phe Glu Gln	Pro		
	1115		1120		1125
Trp Pro Asn Ala	Arg Pro Ala Thr Gln	Ser Gly Arg His Ser	Val		
	1130		1135		1140
Ser Val Glu Val	Gln Met Gln Arg Gln	Arg Gln Glu Glu Arg	Glu		
	1145		1150		1155

Ser Ser Gln Gln Ala Gln Arg Gln Tyr Ser Ser Leu Pro Arg Gln  
 1160 1165 1170  
 Ser Arg Lys Asn Ala Ser Ser Val Ser Gln Asp Ser Trp Glu Gln  
 1175 1180 1185  
 Asn Tyr Ser Pro Gly Glu Gly Phe Gln Ser Ala Lys Glu Asn Pro  
 1190 1195 1200  
 Arg Tyr Ser Ser Tyr Gln Gly Ser Arg Asn Gly Tyr Leu Gly Gly  
 1205 1210 1215  
 His Gly Phe Asn Ala Arg Val Met Leu Glu Thr Gln Glu Leu Leu  
 1220 1225 1230  
 Arg Gln Glu Gln Arg Arg Lys Glu Gln Gln Met Lys Lys Gln Pro  
 1235 1240 1245  
 Pro Ser Glu Gly Pro Ser Asn Tyr Asp Ser Tyr Lys Lys Val Gln  
 1250 1255 1260  
 Asp Pro Ser Tyr Ala Pro Pro Lys Gly Pro Phe Arg Gln Asp Val  
 1265 1270 1275  
 Pro Pro Ser Pro Ser Gln Val Ala Arg Leu Asn Arg Leu Gln Thr  
 1280 1285 1290  
 Pro Glu Lys Gly Arg Pro Phe Tyr Ser  
 1295

<210> 3  
 <211> 433  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7502420CD1

<400> 3  
 Met Glu Ala Cys Ser Pro Cys Arg Ser Glu Thr Tyr Glu Leu Gln  
 1 5 10 15  
 Leu Pro Ala Ser Ile Gly Arg Pro Arg Arg Asn Cys Gly Gln Pro  
 20 25 30  
 Ile Val Pro Glu Gln Gly Lys Pro Pro Pro Ser Gly Ser Val Glu  
 35 40 45  
 Gly Leu Arg Gln Trp Gln Ser Ser His Arg Arg Ala Cys Gly Leu  
 50 55 60  
 Gly Cys Glu Ala Arg Ala Gly Gly Gly Glu Glu Pro Arg Gly Arg  
 65 70 75  
 Ala Ser Ser Val Ala Gly Trp Val Gly Ala Phe Arg Ala Pro Phe  
 80 85 90  
 Ile Glu Ala Ala Val Ala Gly Leu Gly Ala Gly Ser Gly Lys Arg  
 95 100 105  
 Arg Arg Gly Trp Lys Met Pro Val His Ser Arg Gly Asp Lys Lys  
 110 115 120  
 Glu Thr Asn His His Asp Glu Met Glu Val Asp Tyr Ala Glu Asn  
 125 130 135  
 Glu Gly Ser Ser Ser Glu Asp Glu Asp Thr Glu Ser Ser Ser Val  
 140 145 150  
 Ser Glu Asp Gly Asp Ser Ser Glu Met Asp Asp Glu Asp Cys Glu  
 155 160 165  
 Arg Arg Arg Met Glu Cys Leu Asp Glu Met Ser Asn Leu Glu Lys  
 170 175 180  
 Gln Phe Thr Asp Leu Lys Asp Gln Leu Tyr Lys Glu Arg Leu Ser  
 185 190 195  
 Gln Val Asp Ala Lys Leu Gln Glu Val Ile Ala Gly Lys Ala Pro  
 200 205 210  
 Glu Tyr Leu Glu Pro Leu Ala Thr Leu Gln Glu Asn Met Gln Ile  
 215 220 225  
 Arg Thr Lys Val Ala Gly Ile Tyr Arg Glu Leu Cys Leu Glu Ser  
 230 235 240

Val	Lys	Asn	Lys	Tyr	Glu	Cys	Glu	Ile	Gln	Ala	Ser	Arg	Gln	His
				245					250					255
Cys	Glu	Ser	Glu	Lys	Leu	Leu	Leu	Tyr	Asp	Thr	Val	Gln	Ser	Glu
				260					265					270
Leu	Glu	Glu	Lys	Ile	Arg	Arg	Leu	Glu	Glu	Asp	Arg	His	Ser	Ile
				275					280					285
Asp	Ile	Thr	Ser	Glu	Leu	Trp	Asn	Asp	Glu	Leu	Gln	Ser	Arg	Lys
				290					295					300
Lys	Arg	Lys	Asp	Pro	Phe	Ser	Pro	Asp	Lys	Lys	Pro	Val	Val	
				305					310					315
Val	Ser	Gly	Pro	Tyr	Ile	Val	Tyr	Met	Leu	Gln	Asp	Leu	Asp	Ile
				320					325					330
Leu	Glu	Asp	Trp	Thr	Thr	Ile	Arg	Lys	Ala	Met	Ala	Thr	Leu	Gly
				335					340					345
Pro	His	Arg	Val	Lys	Thr	Glu	Pro	Pro	Val	Lys	Leu	Glu	Lys	His
				350					355					360
Leu	His	Ser	Ala	Arg	Ser	Glu	Glu	Gly	Arg	Leu	Tyr	Tyr	Asp	Gly
				365					370					375
Glu	Trp	Tyr	Ile	Arg	Gly	Gln	Thr	Ile	Cys	Ile	Asp	Lys	Lys	Asp
				380					385					390
Glu	Cys	Pro	Thr	Ser	Ala	Val	Ile	Thr	Thr	Ile	Asn	His	Asp	Glu
				395					400					405
Val	Trp	Phe	Lys	Arg	Pro	Asp	Gly	Ser	Lys	Ser	Lys	Leu	Tyr	Ile
				410					415					420
Ser	Gln	Leu	Gln	Lys	Gly	Lys	Tyr	Ser	Ile	Lys	His	Ser		
				425					430					

&lt;210&gt; 4

&lt;211&gt; 199

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506455CD1

&lt;400&gt; 4

Met	Glu	Cys	Arg	Lys	Met	Asn	Gly	Arg	Ala	Asp	Phe	Arg	Glu	Pro
1				5					10					15
Asn	Ala	Glu	Val	Pro	Arg	Pro	Ile	Pro	His	Ile	Gly	Pro	Asp	Tyr
				20					25					30
Ile	Pro	Thr	Glu	Glu	Glu	Arg	Arg	Val	Phe	Ala	Glu	Cys	Asn	Asp
				35					40					45
Glu	Ser	Phe	Trp	Phe	Arg	Ser	Val	Pro	Leu	Ala	Ala	Thr	Ser	Met
				50					55					60
Leu	Ile	Thr	Gln	Gly	Leu	Ile	Ser	Lys	Gly	Ile	Leu	Ser	Ser	His
				65					70					75
Pro	Lys	Tyr	Gly	Ser	Ile	Pro	Lys	Leu	Ile	Leu	Ala	Cys	Ile	Met
				80					85					90
Gly	Tyr	Phe	Ala	Gly	Lys	Leu	Ser	Tyr	Val	Lys	Thr	Cys	Gln	Glu
				95					100					105
Lys	Phe	Lys	Lys	Leu	Glu	Asn	Ser	Pro	Leu	Gly	Glu	Ala	Leu	Arg
				110					115					120
Ser	Gly	Gln	Ala	Arg	Arg	Ser	Ser	Pro	Pro	Gly	His	Tyr	Tyr	Gln
				125					130					135
Lys	Ser	Lys	Tyr	Asp	Ser	Ser	Val	Ser	Gly	Gln	Ser	Ser	Phe	Val
				140					145					150
Thr	Ser	Pro	Ala	Ala	Asp	Asn	Ile	Glu	Met	Leu	Pro	His	Tyr	Glu
				155					160					165
Pro	Ile	Pro	Phe	Ser	Ser	Ser	Met	Asn	Glu	Ser	Ala	Pro	Thr	Gly
				170					175					180
Ile	Thr	Asp	His	Ile	Val	Gln	Val	Lys	Val	Asn	Lys	Tyr	Gly	Asp
				185					190					195

Thr Trp Asp Glu

<210> 5  
 <211> 37  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7506018CD1

<400> 5  
 Met Ser Gly Leu Arg Val Tyr Ser Thr Ser Val Thr Gly Ser Arg  
 1 5 10 15  
 Glu Asp Tyr Glu Leu Phe Val Glu Ala Val Glu Gln Asn Thr Leu  
 20 25 30  
 Gln Glu Phe Leu Lys Leu Ala  
 35

<210> 6  
 <211> 312  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 353017CD1

<400> 6  
 Met Ala Leu Leu Tyr Glu Gly Asp Ile Ser Arg Gly Ile Ala Asp  
 1 5 10 15  
 Gln Glu Ile Ile Ala Gly Gly Asp Val Lys Tyr Thr Thr Trp Met  
 20 25 30  
 Phe Glu Thr Gln Ala Ile Asp Thr Leu Gly Ala His Ser Ser Asp  
 35 40 45  
 Thr Val Glu Asn Ala Glu Lys Ile Pro Glu Leu Ala Arg Gly Asp  
 50 55 60  
 Val Cys Thr Ala Arg Trp Met Phe Glu Thr Arg Pro Leu Asp Ser  
 65 70 75  
 Met Asn Lys Met His Gln Ser Gln Glu Glu Ser Ala Val Thr Ile  
 80 85 90  
 Ser Lys Asp Ile Thr Gly Gly Asp Val Lys Thr Val Arg Tyr Met  
 95 100 105  
 Phe Glu Thr Gln His Leu Asp Gln Leu Gly Gln Leu His Ser Val  
 110 115 120  
 Asp Glu Val His Leu Leu Gln Leu Arg Ser Glu Leu Lys Glu Ile  
 125 130 135  
 Lys Gly Asn Val Lys Arg Ser Ile Lys Cys Phe Glu Thr Gln Pro  
 140 145 150  
 Leu Tyr Val Ile Arg Asp Gly Ser Gly Gln Met Leu Glu Ile Lys  
 155 160 165  
 Thr Val His Arg Glu Asp Val Glu Lys Gly Asp Val Arg Thr Ala  
 170 175 180  
 Arg Trp Met Phe Glu Thr Gln Pro Leu Asp Thr Ile Asn Lys Asp  
 185 190 195  
 Ile Thr Glu Ile Lys Val Val Arg Gly Ile Ser Met Glu Glu Asn  
 200 205 210  
 Val Lys Gly Gly Val Ser Lys Ala Lys Trp Leu Phe Glu Thr Gln  
 215 220 225  
 Pro Leu Glu Lys Ile Lys Glu Ser Glu Val Ile Ile Glu Lys  
 230 235 240  
 Glu Lys Ile Ile Gly Thr Asp Val Ser Arg Lys Cys Trp Met Phe

	245	250	255
Glu Thr Gln Pro	Leu Asp Ile Leu Lys	Glu Val Pro Asp Ala	Asp
	260	265	270
Ser Leu Gln Arg	Glu Glu Ile Ile Gly	Gly Asp Val Gln Thr	Thr
	275	280	285
Lys His Leu Phe	Glu Thr Leu Pro Ile	Glu Ala Leu Lys Asp	Ser
	290	295	300
Pro Asp Ile Gly	Lys Leu Gln Lys Asn	His Cys Leu	
	305	310	

&lt;210&gt; 7

&lt;211&gt; 993

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5137816CD1

&lt;400&gt; 7

Met	Glu	Lys	Gln	Leu	Ser	Pro	Arg	Met	Gln	Asn	Asn	Glu	Glu	Leu
1				5					10					15
Pro	Thr	Tyr	Glu	Glu	Ala	Lys	Val	Gln	Ser	Gln	Tyr	Phe	Arg	Gly
				20					25					30
Gln	Gln	His	Ala	Ser	Val	Gly	Ala	Ala	Phe	Tyr	Val	Thr	Gly	Val
				35					40					45
Thr	Asn	Gln	Lys	Met	Arg	Thr	Glu	Gly	Arg	Pro	Ser	Val	Gln	Arg
				50					55					60
Leu	Asn	Pro	Gly	Lys	Met	His	Gln	Asp	Glu	Gly	Leu	Arg	Asp	Leu
				65					70					75
Lys	Gln	Gly	His	Val	Arg	Ser	Leu	Ser	Glu	Arg	Leu	Met	Gln	Met
				80					85					90
Ser	Leu	Ala	Thr	Ser	Gly	Val	Lys	Ala	His	Pro	Pro	Val	Thr	Ser
				95					100					105
Ala	Pro	Leu	Ser	Pro	Pro	Gln	Pro	Asn	Asp	Leu	Tyr	Lys	Asn	Pro
				110					115					120
Thr	Ser	Ser	Ser	Glu	Phe	Tyr	Lys	Ala	Gln	Gly	Pro	Leu	Pro	Asn
				125					130					135
Gln	His	Ser	Leu	Lys	Gly	Met	Glu	His	Arg	Gly	Pro	Pro	Pro	Glu
				140					145					150
Tyr	Pro	Phe	Lys	Gly	Met	Pro	Pro	Gln	Ser	Val	Val	Cys	Lys	Pro
				155					160					165
Gln	Glu	Pro	Gly	His	Phe	Tyr	Ser	Glu	His	Arg	Leu	Asn	Gln	Pro
				170					175					180
Gly	Arg	Thr	Glu	Gly	Gln	Leu	Met	Arg	Tyr	Gln	His	Pro	Pro	Glu
				185					190					195
Tyr	Gly	Ala	Ala	Arg	Pro	Ala	Gln	Asp	Ile	Ser	Leu	Pro	Leu	Ser
				200					205					210
Ala	Arg	Asn	Ser	Gln	Pro	His	Ser	Pro	Thr	Ser	Ser	Leu	Thr	Ser
				215					220					225
Gly	Gly	Ser	Leu	Pro	Leu	Leu	Gln	Ser	Pro	Pro	Ser	Thr	Arg	Leu
				230					235					240
Ser	Pro	Ala	Arg	His	Pro	Leu	Val	Pro	Asn	Gln	Gly	Asp	His	Ser
				245					250					255
Ala	His	Leu	Pro	Arg	Pro	Gln	Gln	His	Phe	Leu	Pro	Asn	Gln	Ala
				260					265					270
His	Gln	Gly	Asp	His	Tyr	Arg	Leu	Ser	Gln	Pro	Gly	Leu	Ser	Gln
				275					280					285
Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	His	His	His	His	His	His	His
				290					295					300
Gln	Gln	Gln	Gln	Gln	Gln	Gln	Pro	Gln	Gln	Gln	Pro	Gly	Glu	Ala
				305					310					315
Tyr	Ser	Ala	Met	Pro	Arg	Ala	Gln	Pro	Ser	Ser	Ala	Ser	Tyr	Gln



Pro Val Pro Ala	320	325	330
Asp Pro Phe Ala Ile	335	Val Ser Arg Ala Gln	Gln
Met Val Glu Ile	350	340	345
Leu Ser Asp Glu Asn	355	Arg Asn Leu Arg Gln	Glu
Leu Glu Gly Cys	365	370	375
Tyr Glu Lys Val Ala	380	Arg Leu Gln Lys Val	Glu
Thr Glu Ile Gln	395	385	390
Arg Val Ser Glu Ala	410	Tyr Glu Asn Leu Val	Lys
Ser Ser Ser Lys	425	395	400
Arg Glu Ala Leu Glu	440	Lys Ala Met Arg Asn	Lys
Leu Glu Gly Glu	455	400	405
Ile Arg Arg Met His	470	415	420
Arg Glu Arg Leu	485	430	435
Glu Thr Ala Asn Lys	500	Gln Leu Ala Glu Lys	Glu
Tyr Glu Gly Ser	515	445	450
Glu Asp Thr Arg Lys	530	460	465
Ala Lys Asn Lys	545	475	480
Glu Ser Gln Arg Glu	560	490	495
Lys Glu Lys Leu Glu	575	505	510
Val Lys Leu Glu	590	515	520
Glu Glu Leu Lys Lys	605	530	535
Lys Val Glu Lys	620	545	550
Met Gln Gln Ala Leu	635	560	565
Cys Glu Lys Arg	650	580	585
Glu Gln Leu Glu His	665	595	600
Arg Leu Arg Thr Arg	680	610	615
Glu Arg Glu Leu	695	625	630
Glu Ser Leu Arg Ile	710	640	645
Gln Gln Arg Gln Gly	725	655	660
Asn	740	670	675
Cys Gln Pro Thr	755	685	690
Asn Val Ser Glu Tyr	770	700	705
Ala Ala Ala Ala Leu	785	715	720
Met	790	730	735
Glu Leu Leu Arg		745	750
Glu Lys Glu Glu Arg		760	765
Ile Leu Ala Leu Glu		775	780
Ala Leu Glu Ala		790	795
Asp Met Thr Lys			
Trp Glu Gln Lys Tyr			
Leu Glu Glu Asn Val			
Met			
Arg His Phe Ala			
Leu Asp Ala Ala Ala			
Thr Val Ala Ala Gln			
Arg			
Asp Thr Thr Val			
Ile Ser His Ser Pro			
Asn Thr Ser Tyr Asp			
Thr			
Ala Leu Glu Ala			
Arg Ile Gln Lys Glu			
Glu Glu Glu Ile Leu			
Met			
Ala Asn Lys Arg			
Cys Leu Asp Met Glu			
Gly Arg Ile Lys Thr			
Leu			
His Ala Gln Ile			
Ile Glu Lys Asp Ala			
Met Ile Lys Val Leu			
Gln			
Gln Arg Ser Arg			
Lys Glu Pro Ser Lys			
Thr Glu Gln Leu Ser			
Cys			
Met Arg Pro Ala			
Lys Ser Leu Met Ser			
Ile Ser Asn Ala Gly			
Ser			
Gly Leu Leu Ser			
His Ser Ser Thr Leu			
Thr Gly Ser Pro Ile			
Met			
Glu Glu Lys Arg			
Asp Lys Ser Trp			
Lys Gly Ser Leu Gly			
Ile			
Leu Leu Gly Gly			
Asp Tyr Arg Ala Glu			
Tyr Val Pro Ser Thr			
Pro			
Ser Pro Val Pro			
Pro Ser Thr Pro Leu			
Leu Ser Ala His Ser			
Lys			
Thr Gly Ser Arg			
Asp Cys Ser Thr Gln			
Thr Glu Arg Gly Thr			
Glu			
Ser Asn Lys Thr			
Ala Ala Val Ala Pro			
Ile Ser Val Pro Ala			
Pro			

```

Val Ala Ala Ala Ala Thr Ala Ala Ala Ile Thr Ala Thr Ala Ala
      800      805      810
Ile Ile Thr Ser Ile Met Val Ala Ala Ala Pro Val Ala Val Ala
      815      820      825
Ala Ala Ala Ala Pro Ala Ala Ala Ala Ala Pro Ser Pro Ala Thr
      830      835      840
Ala Ala Ala Thr Ala Ala Ala Val Ser Pro Ala Ala Ala Gly Gln
      845      850      855
Ile Pro Ala Ala Ala Ser Val Ala Ser Ala Ala Ala Val Ala Pro
      860      865      870
Ser Ala Ala Ala Ala Ala Ala Val Gln Val Ala Pro Ala Ala Pro
      875      880      885
Ala Pro Val Pro Ala Pro Ala Leu Val Pro Val Pro Ala Pro Ala
      890      895      900
Ala Ala Gln Ala Ser Ala Pro Ala Gln Thr Gln Ala Pro Thr Ser
      905      910      915
Ala Pro Ala Val Ala Pro Thr Pro Ala Pro Thr Pro Thr Pro Ala
      920      925      930
Val Ala Gln Ala Glu Val Pro Ala Ser Pro Ala Thr Gly Pro Gly
      935      940      945
Pro His Arg Leu Ser Ile Pro Ser Leu Thr Cys Asn Pro Asp Lys
      950      955      960
Thr Asp Gly Pro Val Phe His Ser Asn Thr Leu Glu Arg Lys Thr
      965      970      975
Pro Ile Gln Ile Leu Gly Gln Glu Pro Asp Ala Glu Met Val Glu
      980      985      990
Tyr Leu Ile

```

&lt;210&gt; 8

&lt;211&gt; 204

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7502151CD1

&lt;400&gt; 8

```

Met Glu Ala Ala Gly Ser Pro Ala Ala Thr Glu Thr Gly Lys Tyr
  1      5      10      15
Ile Ala Ser Thr Gln Arg Pro Asp Gly Thr Trp Arg Lys Gln Arg
      20      25      30
Arg Val Lys Glu Gly Tyr Val Pro Gln Glu Glu Val Pro Val Tyr
      35      40      45
Glu Asn Lys Tyr Val Lys Phe Phe Lys Ser Lys Pro Glu Leu Pro
      50      55      60
Pro Gly Leu Ser Pro Glu Ala Thr Ala Pro Val Thr Pro Ser Arg
      65      70      75
Pro Glu Gly Gly Glu Pro Gly Leu Ser Lys Thr Ala Lys Arg Asn
      80      85      90
Leu Lys Arg Lys Glu Lys Arg Arg Gln Gln Gln Glu Lys Gly Glu
      95      100      105
Ala Glu Ala Leu Ser Arg Thr Leu Asp Lys Glu Ser Leu Glu Glu
      110      115      120
Thr Ala Gln Leu Pro Ser Ala Pro Gln Gly Ser Arg Ala Ala Pro
      125      130      135
Thr Ala Ala Ser Asp Gln Pro Asp Ser Ala Ala Thr Thr Glu Lys
      140      145      150
Ala Lys Lys Ile Lys Asn Leu Lys Lys Lys Leu Arg Gln Val Glu
      155      160      165
Glu Leu Gln Gln Arg Ile Gln Ala Gly Glu Val Ser Gln Pro Ser
      170      175      180

```

Lys Glu Gln Leu Glu Lys Leu Ala Arg Arg Arg Ala Leu Glu Glu  
 185 190 195  
 Glu Leu Glu Asp Leu Glu Leu Gly Leu  
 200

<210> 9  
 <211> 630  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7505983CD1

<400> 9  
 Met Gly Arg Lys Leu Asp Pro Thr Lys Glu Lys Arg Gly Pro Gly  
 1 5 10 15  
 Arg Lys Ala Arg Lys Gln Lys Gly Ala Glu Thr Glu Leu Val Arg  
 20 25 30  
 Phe Leu Pro Ala Val Ser Asp Glu Asn Ser Lys Arg Leu Ser Ser  
 35 40 45  
 Arg Ala Arg Lys Arg Ala Ala Lys Arg Arg Leu Gly Ser Val Glu  
 50 55 60  
 Ala Pro Lys Thr Asn Lys Ser Pro Glu Ala Lys Pro Leu Pro Gly  
 65 70 75  
 Lys Leu Pro Lys Gly Ile Ser Ala Gly Ala Val Gln Thr Ala Gly  
 80 85 90  
 Lys Lys Gly Pro Gln Ser Leu Phe Asn Ala Pro Arg Gly Lys Lys  
 95 100 105  
 Arg Pro Ala Pro Gly Ser Asp Glu Glu Glu Glu Phe Leu Glu  
 110 115 120  
 Ala Asn Glu Val Pro Arg Pro Val Thr Leu Arg Thr Asn Thr Leu  
 125 130 135  
 Lys Thr Arg Arg Arg Asp Leu Ala Gln Ala Leu Ile Asn Arg Gly  
 140 145 150  
 Val Asn Leu Asp Pro Leu Gly Lys Trp Ser Lys Thr Gly Leu Val  
 155 160 165  
 Val Tyr Asp Ser Ser Val Pro Ile Gly Ala Thr Pro Glu Tyr Leu  
 170 175 180  
 Ala Gly His Tyr Met Leu Gln Gly Ala Ser Ser Met Leu Pro Val  
 185 190 195  
 Met Ala Leu Ala Pro Gln Glu His Glu Arg Ile Leu Asp Met Cys  
 200 205 210  
 Cys Ala Pro Gly Gly Lys Thr Ser Tyr Met Ala Gln Leu Met Lys  
 215 220 225  
 Asn Thr Gly Val Ile Leu Ala Asn Asp Ala Asn Ala Glu Arg Leu  
 230 235 240  
 Lys Ser Val Val Gly Asn Leu His Arg Leu Gly Val Thr Asn Thr  
 245 250 255  
 Ile Ile Ser His Tyr Asp Gly Arg Gln Phe Pro Lys Val Val Gly  
 260 265 270  
 Gly Phe Asp Arg Val Leu Leu Asp Ala Pro Cys Ser Gly Thr Gly  
 275 280 285  
 Val Ile Ser Lys Asp Pro Ala Val Lys Thr Asn Lys Asp Glu Lys  
 290 295 300  
 Asp Ile Leu Arg Cys Ala His Leu Gln Lys Glu Leu Leu Leu Ser  
 305 310 315  
 Ala Ile Asp Ser Val Asn Ala Thr Ser Lys Thr Gly Gly Tyr Leu  
 320 325 330  
 Val Tyr Cys Thr Cys Ser Ile Thr Val Glu Glu Asn Glu Trp Val  
 335 340 345  
 Val Asp Tyr Ala Leu Lys Lys Arg Asn Val Arg Leu Val Pro Thr  
 350 355 360

Gly Leu Asp Phe Gly Gln Glu Gly Phe Thr Arg Phe Arg Glu Arg  
 365 370 375  
 Arg Phe His Pro Ser Leu Arg Ser Thr Arg Arg Phe Tyr Pro His  
 380 385 390  
 Thr His Asn Met Asp Gly Phe Phe Ile Ala Lys Phe Lys Lys Phe  
 395 400 405  
 Ser Asn Ser Ile Pro Gln Ser Gln Thr Gly Asn Ser Glu Thr Ala  
 410 415 420  
 Thr Pro Thr Asn Val Asp Leu Pro Gln Val Ile Pro Lys Ser Glu  
 425 430 435  
 Asn Ser Ser Gln Pro Ala Lys Lys Ala Lys Gly Ala Ala Lys Thr  
 440 445 450  
 Lys Gln Gln Leu Gln Lys Gln Gln His Pro Lys Lys Ala Ser Phe  
 455 460 465  
 Gln Lys Leu Asn Gly Ile Ser Lys Gly Ala Asp Ser Glu Leu Ser  
 470 475 480  
 Thr Val Pro Ser Val Thr Lys Thr Gln Ala Ser Ser Ser Phe Gln  
 485 490 495  
 Asp Ser Ser Gln Pro Ala Gly Lys Ala Glu Gly Ile Arg Glu Pro  
 500 505 510  
 Lys Val Thr Gly Lys Leu Lys Gln Arg Ser Pro Lys Leu Gln Ser  
 515 520 525  
 Ser Lys Lys Val Ala Phe Leu Arg Gln Asn Ala Pro Pro Lys Gly  
 530 535 540  
 Thr Asp Thr Gln Thr Pro Ala Val Leu Ser Pro Ser Lys Thr Gln  
 545 550 555  
 Ala Thr Leu Lys Pro Lys Asp His His Gln Pro Leu Gly Arg Ala  
 560 565 570  
 Lys Gly Val Glu Lys Gln Gln Leu Pro Glu Gln Pro Phe Glu Lys  
 575 580 585  
 Ala Ala Phe Gln Lys Gln Asn Asp Thr Pro Lys Gly Pro Gln Pro  
 590 595 600  
 Pro Thr Val Ser Pro Ile Arg Ser Ser Arg Pro Pro Pro Ala Lys  
 605 610 615  
 Arg Lys Lys Ser Gln Ser Arg Gly Asn Ser Gln Leu Leu Leu Ser  
 620 625 630

<210> 10  
 <211> 129  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7505986CD1

<400> 10  
 Met Gly Arg Lys Leu Asp Pro Thr Lys Glu Lys Arg Gly Pro Gly  
 1 5 10 15  
 Arg Lys Ala Arg Lys Gln Lys Gly Ala Glu Thr Glu Leu Val Arg  
 20 25 30  
 Phe Leu Pro Ala Val Ser Asp Glu Asn Ser Lys Arg Leu Ser Ser  
 35 40 45  
 Arg Ala Arg Lys Arg Ala Ala Lys Arg Arg Leu Gly Ser Val Glu  
 50 55 60  
 Ala Pro Lys Thr Asn Lys Ser Pro Glu Ala Lys Pro Leu Pro Gly  
 65 70 75  
 Lys Leu Pro Lys Gly Ile Ser Ala Gly Ala Val Gln Thr Ala Gly  
 80 85 90  
 Lys Lys Gly Pro Gln Ser Leu Phe Asn Ala Pro Arg Gly Lys Lys  
 95 100 105  
 Arg Pro Leu Ser Pro Gly Trp Asn Ser Cys Leu Cys Glu Asp Ala

110  
 Phe Ser Thr Val His Thr His Glu Ile  
 125

<210> 11  
 <211> 607  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3231075CD1

<400> 11  
 Met Val Thr Asn Lys Met Thr Ala Ala Phe Arg Asn Pro Ser Gly  
 1 5 10 15  
 Lys Gln Val Ala Thr Asp Lys Val Ala Glu Lys Leu Ser Ser Thr  
 20 25 30  
 Leu Ser Trp Val Lys Asn Thr Val Ser His Thr Val Ser Gln Met  
 35 40 45  
 Ala Ser Gln Val Ala Ser Pro Ser Thr Ser Leu His Thr Thr Ser  
 50 55 60  
 Ser Ser Thr Thr Leu Ser Thr Pro Ala Leu Ser Pro Ser Ser Pro  
 65 70 75  
 Ser Gln Leu Ser Pro Asp Val Leu Glu Leu Leu Ala Lys Leu Glu  
 80 85 90  
 Glu Gln Asn Ile Leu Leu Glu Thr Asp Ser Lys Ser Leu Arg Ser  
 95 100 105  
 Val Asn Gly Ser Arg Arg Asn Ser Gly Ser Ser Leu Val Ser Ser  
 110 115 120  
 Ser Ser Ala Ser Ser Asn Leu Ser His Leu Glu Glu Asp Ser Trp  
 125 130 135  
 Ile Leu Trp Gly Arg Ile Val Asn Glu Trp Glu Asp Val Arg Lys  
 140 145 150  
 Lys Lys Glu Lys Gln Val Lys Glu Leu Val His Lys Gly Ile Pro  
 155 160 165  
 His His Phe Arg Ala Ile Val Trp Gln Leu Leu Cys Ser Ala Gln  
 170 175 180  
 Ser Met Pro Ile Lys Asp Gln Tyr Ser Glu Leu Leu Lys Met Thr  
 185 190 195  
 Ser Pro Cys Glu Lys Leu Ile Arg Arg Asp Ile Ala Arg Thr Tyr  
 200 205 210  
 Pro Glu His Asn Phe Phe Lys Glu Lys Asp Ser Leu Gly Gln Glu  
 215 220 225  
 Val Leu Phe Asn Val Met Lys Ala Tyr Ser Leu Val Asp Arg Glu  
 230 235 240  
 Val Gly Tyr Cys Gln Gly Ser Ala Phe Ile Val Gly Leu Leu Leu  
 245 250 255  
 Met Gln Met Pro Glu Glu Glu Ala Phe Cys Val Phe Val Lys Leu  
 260 265 270  
 Met Gln Asp Tyr Arg Leu Arg Glu Leu Phe Lys Pro Ser Met Ala  
 275 280 285  
 Glu Leu Gly Leu Cys Met Tyr Gln Phe Glu Cys Met Ile Gln Glu  
 290 295 300  
 His Leu Pro Glu Leu Phe Val His Phe Gln Ser Gln Ser Phe His  
 305 310 315  
 Thr Ser Met Tyr Ala Ser Ser Trp Phe Leu Thr Ile Phe Leu Thr  
 320 325 330  
 Thr Phe Pro Leu Pro Val Ala Thr Arg Ile Phe Asp Ile Phe Met  
 335 340 345  
 Ser Glu Gly Leu Glu Ile Val Phe Arg Val Gly Leu Ala Leu Leu  
 350 355 360  
 Gln Met Asn Gln Ala Glu Leu Met Gln Leu Asp Met Glu Gly Met

Leu	Gln	His	Phe	365	Gln	Lys	Val	Ile	Pro	370	His	Gln	Phe	Asp	Gly	Val	375
Pro	Asp	Lys	Leu	380	Ile	Gln	Ala	Ala	Tyr	385	Gln	Val	Lys	Tyr	Asn	Ser	390
Lys	Lys	Met	Lys	395	Lys	Leu	Glu	Lys	Glu	400	Tyr	Thr	Thr	Ile	Lys	Thr	405
Lys	Glu	Met	Glu	410	Glu	Gln	Val	Glu	Ile	415	Lys	Arg	Leu	Arg	Thr	Glu	420
Asn	Arg	Leu	Leu	425	Lys	Gln	Arg	Ile	Glu	430	Thr	Leu	Glu	Lys	His	Lys	435
Cys	Ser	Ser	Asn	440	Tyr	Asn	Glu	Asp	Phe	445	Val	Leu	Gln	Leu	Glu	Lys	450
Glu	Leu	Val	Gln	455	Ala	Arg	Leu	Ser	Glu	460	Ala	Glu	Ser	Gln	Cys	Ala	465
Leu	Lys	Glu	Met	470	Gln	Asp	Lys	Val	Leu	475	Asp	Ile	Glu	Lys	Arg	Asn	480
Asn	Ser	Leu	Pro	485	Asp	Glu	Asn	Asn	Ile	490	Ala	Arg	Leu	Gln	Glu	Glu	495
Leu	Ile	Ala	Val	500	Lys	Leu	Arg	Glu	Ala	505	Glu	Ala	Ile	Met	Gly	Leu	510
Lys	Glu	Leu	Arg	515	Gln	Gln	Val	Lys	Asp	520	Leu	Glu	Glu	His	Trp	Gln	525
Arg	His	Leu	Ala	530	Arg	Thr	Thr	Gly	Arg	535	Trp	Lys	Asp	Pro	Pro	Lys	540
Lys	Asn	Ala	Met	545	Asn	Glu	Leu	Gln	Asp	550	Glu	Leu	Met	Thr	Ile	Arg	555
Leu	Arg	Glu	Ala	560	Glu	Thr	Gln	Ala	Glu	565	Ile	Arg	Glu	Ile	Lys	Gln	570
Arg	Met	Met	Glu	575	Met	Glu	Thr	Gln	Leu	580	Val	Glu	Ser	Ala	Asp	Val	585
Glu	Pro	Ala	Tyr	590	Met	Lys	Gly			595							600
				605													

<210> 12  
 <211> 105  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7503516CD1

Met	Gly	Gly	Ala	Gly	Ile	Leu	Leu	Leu	Leu	Leu	Ala	Gly	Ala	Gly			
1				5						10				15			
Val	Val	Val	Ala	Trp	Arg	Pro	Pro	Lys	Gly	Lys	Cys	Pro	Leu	Arg			
				20					25					30			
Cys	Ser	Cys	Ser	Lys	Asp	Ser	Ala	Leu	Cys	Glu	Gly	Ser	Pro	Asp			
				35					40					45			
Leu	Pro	Val	Ser	Phe	Ser	Pro	Thr	Leu	Leu	Ser	Leu	Ser	Leu	Val			
				50					55					60			
Arg	Thr	Gly	Val	Thr	Gln	Leu	Lys	Ala	Gly	Ser	Phe	Leu	Arg	Ile			
				65					70					75			
Pro	Ser	Leu	His	Leu	Leu	Leu	Phe	Thr	Ser	Asn	Ser	Phe	Ser	Val			
				80					85					90			
Ile	Glu	Asp	Asp	Ala	Phe	Ala	Gly	Leu	Ser	His	Leu	Gln	Tyr	Leu			
				95					100					105			

<210> 13  
 <211> 552  
 <212> PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506179CD1

&lt;400&gt; 13

```

Met Ser Asn Glu Val Glu Thr Ser Ala Thr Asn Gly Gln Pro Asp
 1          5          10          15
Gln Gln Ala Ala Pro Lys Ala Pro Ser Lys Lys Glu Lys Lys Lys
          20          25          30
Gly Pro Glu Lys Thr Asp Glu Tyr Leu Leu Ala Arg Phe Lys Gly
          35          40          45
Asp Gly Val Lys Tyr Lys Ala Lys Leu Ile Gly Ile Asp Asp Val
          50          55          60
Pro Asp Ala Arg Gly Asp Lys Met Ser Gln Asp Ser Met Met Lys
          65          70          75
Leu Lys Gly Met Ala Ala Ala Gly Arg Ser Gln Gly Gln His Lys
          80          85          90
Gln Arg Ile Trp Val Asn Ile Ser Leu Ser Gly Ile Lys Ile Ile
          95          100          105
Asp Glu Lys Thr Gly Val Ile Glu His Glu His Pro Val Asn Lys
          110          115          120
Ile Ser Phe Ile Ala Arg Asp Val Thr Asp Asn Arg Ala Phe Gly
          125          130          135
Tyr Val Cys Gly Gly Glu Gly Gln His Gln Phe Phe Ala Ile Lys
          140          145          150
Thr Gly Gln Gln Ala Glu Pro Leu Val Val Asp Leu Lys Asp Leu
          155          160          165
Phe Gln Val Ile Tyr Asn Val Lys Lys Lys Glu Glu Glu Lys Lys
          170          175          180
Lys Ile Glu Glu Ala Ser Lys Ala Val Glu Asn Gly Ser Glu Ala
          185          190          195
Leu Met Ile Leu Asp Asp Gln Thr Asn Lys Leu Lys Ser Gly Val
          200          205          210
Asp Gln Met Asp Leu Phe Gly Asp Met Ser Thr Pro Pro Asp Leu
          215          220          225
Asn Ser Pro Thr Ser Ser Ala Asn Asp Leu Leu Ala Ser Asp Ile
          230          235          240
Phe Ala Pro Pro Val Ser Glu Pro Ser Gly Gln Ala Ser Pro Thr
          245          250          255
Gly Gln Pro Thr Ala Leu Gln Pro Asn Pro Leu Asp Leu Phe Lys
          260          265          270
Thr Ser Ala Pro Ala Pro Val Gly Pro Leu Val Gly Leu Gly Gly
          275          280          285
Val Thr Val Thr Leu Pro Gln Ala Gly Pro Trp Asn Thr Ala Ser
          290          295          300
Leu Val Phe Asn Gln Ser Pro Ser Met Ala Pro Gly Ala Met Met
          305          310          315
Gly Gly Gln Pro Ser Gly Phe Ser Gln Pro Val Ile Phe Gly Thr
          320          325          330
Ser Pro Ala Val Ser Gly Trp Asn Gln Pro Ser Pro Phe Ala Ala
          335          340          345
Ser Thr Pro Pro Pro Val Pro Val Val Trp Gly Pro Ser Ala Ser
          350          355          360
Val Ala Pro Asn Ala Trp Ser Thr Thr Ser Pro Leu Gly Asn Pro
          365          370          375
Phe Gln Ser Asn Ile Phe Pro Ala Pro Ala Val Ser Thr Gln Pro
          380          385          390
Pro Ser Met His Ser Ser Leu Leu Val Thr Pro Pro Gln Pro Pro
          395          400          405
Pro Arg Ala Gly Pro Pro Lys Asp Ile Ser Ser Asp Ala Phe Thr
          410          415          420

```

```

Ala Leu Asp Pro Leu Gly Asp Lys Glu Ile Lys Asp Val Lys Glu
      425      430      435
Met Phe Lys Asp Phe Gln Leu Arg Gln Pro Pro Ala Val Pro Ala
      440      445      450
Arg Lys Gly Glu Gln Thr Ser Ser Gly Thr Leu Ser Ala Phe Ala
      455      460      465
Ser Tyr Phe Asn Ser Lys Val Gly Ile Pro Gln Glu Asn Ala Asp
      470      475      480
His Asp Asp Phe Asp Ala Asn Gln Leu Leu Asn Lys Ile Asn Glu
      485      490      495
Pro Pro Lys Pro Ala Pro Arg Gln Val Ser Leu Pro Val Thr Lys
      500      505      510
Ser Thr Asp Asn Ala Phe Glu Asn Pro Phe Phe Lys Asp Ser Phe
      515      520      525
Gly Ser Ser Gln Ala Ser Val Ala Ser Ser Gln Pro Val Ser Ser
      530      535      540
Glu Met Tyr Arg Asp Pro Phe Gly Asn Pro Phe Ala
      545      550

```

<210> 14  
 <211> 54  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1938744CD1

```

<400> 14
Met Ser Ala Asn Arg Arg Trp Trp Val Pro Pro Asp Asp Glu Asp
  1      5      10      15
Cys Val Ser Glu Lys Leu Leu Arg Lys Thr Arg Glu Ser Pro Leu
      20      25      30
Val Pro Ile Gly Ala Val Tyr Thr Met Tyr Ser Asp Tyr Val Lys
      35      40      45
Arg Met Ala Gln Asp Ala Gly Glu Lys
      50

```

<210> 15  
 <211> 479  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5557436CD1

```

<400> 15
Met Ser Ser Pro Leu Ala Ser Leu Ser Lys Thr Arg Lys Val Pro
  1      5      10      15
Leu Pro Ser Glu Pro Met Asn Pro Gly Arg Arg Gly Ile Arg Ile
      20      25      30
Tyr Gly Asp Val Pro Ala Ser His Asp Ser Glu Leu Met Ala Phe
      35      40      45
Met Thr Arg Lys Leu Trp Asp Leu Glu Gln Gln Val Lys Ala Gln
      50      55      60
Thr Asp Glu Ile Leu Ser Lys Asp Gln Lys Ile Ala Ala Leu Glu
      65      70      75
Asp Leu Val Gln Thr Leu Arg Pro His Pro Ala Glu Ala Thr Leu
      80      85      90
Gln Arg Gln Glu Glu Leu Glu Thr Met Cys Val Gln Leu Gln Arg
      95      100     105
Gln Val Arg Glu Met Glu Arg Phe Leu Ser Asp Tyr Gly Leu Gln

```



	110		115		120
Trp Val Gly Glu	Pro Met Asp Gln Glu	Asp Ser Glu Ser Lys Thr			
	125		130		135
Val Ser Glu His	Gly Glu Arg Asp Trp	Met Thr Ala Lys Lys Phe			
	140		145		150
Trp Lys Pro Gly	Asp Ser Leu Ala Pro	Pro Glu Val Asp Phe Asp			
	155		160		165
Arg Leu Leu Ala	Ser Leu Gln Asp Leu	Ser Glu Leu Val Val Glu			
	170		175		180
Gly Asp Thr Gln	Val Thr Pro Val Pro	Gly Gly Ala Arg Leu Arg			
	185		190		195
Thr Leu Glu Pro	Ile Pro Leu Lys Leu	Tyr Arg Asn Gly Ile Met			
	200		205		210
Met Phe Asp Gly	Pro Phe Gln Pro Phe	Tyr Asp Pro Ser Thr Gln			
	215		220		225
Arg Cys Leu Arg	Asp Ile Leu Asp Gly	Phe Phe Pro Ser Glu Leu			
	230		235		240
Gln Arg Leu Tyr	Pro Asn Gly Val Pro	Phe Lys Val Ser Asp Leu			
	245		250		255
Arg Asn Gln Val	Tyr Leu Glu Asp Gly	Leu Asp Pro Phe Pro Gly			
	260		265		270
Glu Gly Arg Val	Val Gly Arg Gln Arg	Met His Lys Ala Leu Asp			
	275		280		285
Arg Val Glu Glu	His Pro Gly Ser Arg	Met Thr Ala Glu Lys Phe			
	290		295		300
Leu Asn Arg Leu	Pro Lys Phe Val Ile	Arg Gln Gly Glu Val Ile			
	305		310		315
Asp Ile Arg Gly	Pro Ile Arg Asp Thr	Leu Gln Asn Cys Cys Pro			
	320		325		330
Leu Pro Ala Arg	Ile Gln Glu Ile Val	Val Glu Thr Pro Thr Leu			
	335		340		345
Ala Ala Glu Arg	Glu Arg Ser Gln Glu	Ser Pro Asn Thr Pro Ala			
	350		355		360
Pro Pro Leu Ser	Met Leu Arg Ile Lys	Ser Glu Asn Gly Glu Gln			
	365		370		375
Ala Phe Leu Leu	Met Met Gln Pro Asp	Asn Thr Ile Gly Asp Val			
	380		385		390
Arg Ala Leu Leu	Ala Gln Ala Arg Val	Met Asp Ala Ser Ala Phe			
	395		400		405
Glu Ile Phe Ser	Thr Phe Pro Pro Thr	Leu Tyr Gln Asp Asp Thr			
	410		415		420
Leu Thr Leu Gln	Ala Ala Gly Leu Val	Pro Lys Ala Ala Leu Leu			
	425		430		435
Leu Arg Ala Arg	Arg Ala Pro Lys Ser	Ser Leu Lys Phe Ser Pro			
	440		445		450
Gly Pro Cys Pro	Gly Pro Gly Pro Gly	Pro Ser Pro Gly Pro Gly			
	455		460		465
Pro Gly Ser Ser	Pro Cys Pro Gly Pro	Ser Pro Ser Pro Gln			
	470		475		

&lt;210&gt; 16

&lt;211&gt; 295

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506178CD1

&lt;400&gt; 16

Met Lys Lys Glu His Val Leu His Cys Gln Phe Ser Ala Trp Tyr
1 5 10 15
Pro Phe Phe Arg Gly Val Thr Ile Lys Ser Val Ile Leu Pro Leu

```

      . 20      25      30
Pro Gln Asn Val Lys Asp Tyr Leu Leu Asp Asp Gly Thr Leu Val
      35      40      45
Val Ser Gly Arg Asp Asp Pro Pro Thr His Ser Gln Pro Asp Ser
      50      55      60
Asp Asp Glu Ala Glu Glu Ile Gln Trp Ser Asp Asp Glu Asn Thr
      65      70      75
Ala Thr Leu Thr Ala Pro Glu Phe Pro Glu Phe Ala Thr Lys Val
      80      85      90
Gln Glu Ala Ile Asn Ser Leu Gly Gly Ser Val Phe Pro Lys Leu
      95     100     105
Asn Trp Ser Ala Pro Arg Asp Ala Tyr Trp Ile Ala Met Asn Ser
     110     115     120
Ser Leu Lys Cys Lys Thr Leu Ser Asp Ile Phe Leu Leu Phe Lys
     125     130     135
Ser Ser Asp Phe Ile Thr Arg Asp Phe Thr Gln Pro Phe Ile His
     140     145     150
Cys Thr Asp Asp Ser Pro Asp Pro Cys Ile Glu Tyr Glu Leu Val
     155     160     165
Leu Arg Lys Trp Cys Glu Leu Ile Pro Gly Ala Glu Phe Arg Cys
     170     175     180
Phe Val Lys Glu Asn Lys Leu Ile Val Val Phe Asp Ile Tyr Arg
     185     190     195
Asp Ser Arg Gly Lys Val Trp Leu Ile Asp Phe Asn Pro Phe Gly
     200     205     210
Glu Val Thr Asp Ser Leu Leu Phe Thr Trp Glu Glu Leu Ile Ser
     215     220     225
Glu Asn Asn Leu Asn Gly Asp Phe Ser Glu Val Asp Ala Gln Glu
     230     235     240
Gln Asp Ser Pro Ala Phe Arg Cys Thr Asn Ser Glu Val Thr Val
     245     250     255
Gln Pro Ser Pro Tyr Leu Ser Tyr Arg Leu Pro Lys Asp Phe Val
     260     265     270
Asp Leu Ser Thr Gly Glu Asp Ala His Lys Leu Ile Asp Phe Leu
     275     280     285
Lys Leu Lys Arg Asn Gln Gln Glu Asp Asp
     290     295

```

&lt;210&gt; 17

&lt;211&gt; 214

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506235CD1

&lt;400&gt; 17

```

Met Cys Ile Ile Phe Phe Lys Phe Asp Pro Arg Pro Val Ser Lys
  1      5      10      15
Asn Ala Tyr Arg Leu Ile Leu Ala Ala Asn Arg Asp Glu Phe Tyr
     20      25      30
Ser Arg Pro Ser Lys Leu Ala Asp Phe Trp Gly Asn Asn Asn Glu
     35      40      45
Ile Leu Ser Gly Leu Asp Met Glu Glu Gly Lys Glu Gly Gly Thr
     50      55      60
Trp Leu Gly Ile Ser Thr Arg Gly Lys Leu Ala Ala Leu Thr Asn
     65      70      75
Tyr Leu Gln Pro Gln Leu Asp Trp Gln Ala Arg Gly Arg Gly Thr
     80      85      90
Tyr Gly Leu Ser Asn Ala Leu Leu Glu Thr Pro Trp Arg Lys Leu
     95     100     105
Cys Phe Gly Lys Gln Leu Phe Leu Glu Ala Val Glu Arg Ser Gln

```

	110		115		120
Ala Leu Pro Lys	Asp Val Leu Ile Ala	Ser Leu Leu Asp Val Leu			
	125		130		135
Asn Asn Glu Glu	Ala Gln Leu Pro Asp	Pro Ala Ile Glu Asp Gln			
	140		145		150
Gly Gly Glu Tyr	Val Gln Pro Met Leu	Ser Lys Tyr Ala Ala Val			
	155		160		165
Cys Val Arg Cys	Pro Gly Tyr Gly Thr	Arg Thr Asn Thr Ile Ile			
	170		175		180
Leu Val Asp Ala	Asp Gly His Val Thr	Phe Thr Glu Arg Ser Met			
	185		190		195
Met Asp Lys Asp	Leu Ser His Trp Glu	Thr Arg Thr Tyr Glu Phe			
	200		205		210
Thr Leu Gln Ser					

&lt;210&gt; 18

&lt;211&gt; 1596

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1302184CD1

&lt;400&gt; 18

Met Glu Ser Arg Glu Thr Leu Ser Ser Ser Arg Gln Arg Gly Gly		
1	5	10
Glu Ser Asp Phe Leu Pro Val Ser Ser Ala Lys Pro Pro Ala Ala		15
	20	25
Pro Gly Cys Ala Gly Glu Pro Leu Leu Ser Thr Pro Gly Pro Gly		30
	35	40
Lys Gly Ile Pro Val Gly Gly Glu Arg Met Glu Pro Glu Glu Glu		45
	50	55
Asp Glu Leu Gly Ser Gly Arg Asp Val Asp Ser Asn Ser Asn Ala		60
	65	70
Asp Ser Glu Lys Trp Val Ala Gly Asp Gly Leu Glu Glu Gln Glu		75
	80	85
Phe Ser Ile Lys Glu Ala Asn Phe Thr Glu Gly Ser Leu Lys Leu		90
	95	100
Lys Ile Gln Thr Thr Lys Arg Ala Lys Lys Pro Pro Lys Asn Leu		105
	110	115
Glu Asn Tyr Ile Cys Pro Pro Glu Ile Lys Ile Thr Ile Lys Gln		120
	125	130
Ser Gly Asp Gln Lys Val Ser Arg Ala Gly Lys Asn Ser Lys Ala		135
	140	145
Thr Lys Glu Glu Glu Arg Ser His Ser Lys Lys Lys Leu Leu Thr		150
	155	160
Ala Ser Asp Leu Ala Ala Ser Asp Leu Lys Gly Phe Gln Pro Gln		165
	170	175
Ala Tyr Glu Arg Pro Gln Lys His Ser Thr Leu His Tyr Asp Thr		180
	185	190
Gly Leu Pro Gln Asp Phe Thr Gly Asp Thr Leu Lys Pro Lys His		195
	200	205
Gln Gln Lys Ser Ser Ser Gln Asn His Met Asp Trp Ser Thr Asn		210
	215	220
Ser Asp Ser Gly Pro Val Thr Gln Asn Cys Phe Ile Ser Pro Glu		225
	230	235
Ser Gly Arg Glu Thr Ala Ser Thr Ser Lys Ile Pro Ala Leu Glu		240
	245	250
Pro Val Ala Ser Phe Ala Lys Ala Gln Gly Lys Lys Gly Ser Ala		255
	260	265
Gly Asn Thr Trp Ser Gln Leu Ser Asn Asn Asn Lys Asp Leu Leu		270

	275		280		285
Leu Gly Gly Val	Ala Pro Ser Pro Ser	Ser His Ser Ser Pro	Ala		
	290		295		300
Pro Pro Ser Ser	Ser Ala Glu Cys Asn	Gly Leu Gln Pro Leu	Val		
	305		310		315
Asp Gln Asp Gly	Gly Gly Thr Lys Glu	Pro Pro Glu Pro Pro	Thr		
	320		325		330
Val Gly Ser Lys	Lys Lys Ser Ser Lys	Lys Asp Val Ile Ser	Gln		
	335		340		345
Thr Ile Pro Asn	Pro Asp Leu Asp Trp	Val Lys Asn Ala Gln	Lys		
	350		355		360
Ala Phe Asp Asn	Thr Glu Gly Lys Arg	Gly Gly Tyr Ser Ala	Asp		
	365		370		375
Ser Ala Gln Glu	Ala Ser Pro Ala Arg	Gln Asn Val Ser Ser	Ala		
	380		385		390
Ser Asn Pro Glu	Asn Asp Ser Ser His	Val Arg Ile Thr Ile	Pro		
	395		400		405
Ile Lys Ala Pro	Ser Leu Asp Pro Thr	Asn His Lys Arg Lys	Lys		
	410		415		420
Arg Gln Ser Ile	Lys Ala Val Val Glu	Lys Ile Met Pro Glu	Lys		
	425		430		435
Ala Leu Ala Ser	Gly Ile Thr Met Ser	Ser Glu Val Val Asn	Arg		
	440		445		450
Ile Leu Ser Asn	Ser Glu Gly Asn Lys	Lys Asp Pro Arg Val	Pro		
	455		460		465
Lys Leu Ser Lys	Met Ile Glu Asn Glu	Ser Pro Ser Val Gly	Leu		
	470		475		480
Glu Thr Gly Gly	Asn Ala Glu Lys Val	Ile Pro Gly Gly Val	Ser		
	485		490		495
Lys Pro Arg Lys	Pro Pro Met Val Met	Thr Pro Pro Thr Cys	Thr		
	500		505		510
Asp His Ser Pro	Ser Arg Lys Leu Pro	Glu Ile Gln His Pro	Lys		
	515		520		525
Phe Ala Ala Lys	Arg Arg Trp Thr Cys	Ser Lys Pro Lys Pro	Ser		
	530		535		540
Thr Met Leu Arg	Glu Ala Val Met Ala	Thr Ser Asp Lys Leu	Met		
	545		550		555
Leu Glu Pro Pro	Ser Ala Tyr Pro Ile	Thr Pro Ser Ser Pro	Leu		
	560		565		570
Tyr Thr Asn Thr	Asp Ser Leu Thr Val	Ile Thr Pro Val Lys	Lys		
	575		580		585
Lys Arg Gly Arg	Pro Lys Lys Gln Pro	Leu Leu Thr Val Glu	Thr		
	590		595		600
Ile His Glu Gly	Thr Ser Thr Ser Pro	Val Ser Pro Ile Ser	Arg		
	605		610		615
Glu Phe Pro Gly	Thr Lys Lys Arg Lys	Arg Arg Arg Asn Leu	Ala		
	620		625		630
Lys Leu Ala Gln	Leu Val Pro Gly Glu	Asp Lys Pro Met Ser	Glu		
	635		640		645
Met Lys Phe His	Lys Lys Val Gly Lys	Leu Gly Val Leu Asp	Lys		
	650		655		660
Lys Thr Ile Lys	Thr Ile Asn Lys Met	Lys Thr Leu Lys Arg	Lys		
	665		670		675
Asn Ile Leu Asn	Gln Ile Leu Ser Cys	Ser Ser Ser Val Ala	Leu		
	680		685		690
Lys Ala Lys Ala	Pro Pro Glu Thr Ser	Pro Gly Ala Ala Ala	Ile		
	695		700		705
Glu Ser Lys Leu	Gly Lys Gln Ile Asn	Val Ser Lys Arg Gly	Thr		
	710		715		720
Ile Tyr Ile Gly	Lys Lys Arg Gly Arg	Lys Pro Arg Ala Glu	Leu		
	725		730		735
Pro Pro Pro Ser	Glu Glu Pro Lys Thr	Ala Ile Lys His Pro	Arg		
	740		745		750

Pro Val Ser Ser	Gln	Pro Asp Val	Pro Ala	Val	Pro Ser Asn	Phe
	755		760			765
Gln Ser Leu Val	Ala	Ser Ser Pro	Ala Ala	Met His	Pro Leu	Ser
	770		775			780
Thr Gln Leu Gly	Gly	Ser Asn Gly	Asn Leu	Ser Pro	Ala Ser	Thr
	785		790			795
Glu Thr Asn Phe	Ser	Glu Leu Lys	Thr Met	Pro Asn	Leu Gln	Pro
	800		805			810
Ile Ser Ala Leu	Pro	Thr Lys Thr	Gln Lys	Gly Ile	His Ser	Gly
	815		820			825
Thr Trp Lys Leu	Ser	Pro Pro Arg	Leu Met	Ala Asn	Ser Pro	Ser
	830		835			840
His Leu Cys Glu	Ile	Gly Ser Leu	Lys Glu	Ile Thr	Leu Ser	Pro
	845		850			855
Val Ser Glu Ser	His	Ser Glu Glu	Thr Ile	Pro Ser	Asp Ser	Gly
	860		865			870
Ile Gly Thr Asp	Asn	Asn Ser Thr	Ser Asp	Gln Ala	Glu Lys	Ser
	875		880			885
Ser Glu Ser Arg	Arg	Arg Tyr Ser	Phe Asp	Phe Cys	Ser Leu	Asp
	890		895			900
Asn Pro Glu Ala	Ile	Pro Ser Asp	Thr Ser	Thr Lys	Asn Arg	His
	905		910			915
Gly His Arg Gln	Lys	His Leu Ile	Val Asp	Asn Phe	Leu Ala	His
	920		925			930
Glu Ser Leu Lys	Lys	Pro Lys His	Lys Arg	Lys Arg	Lys Ser	Leu
	935		940			945
Gln Asn Arg Asp	Asp	Leu Gln Phe	Leu Ala	Asp Leu	Glu Glu	Leu
	950		955			960
Ile Thr Lys Phe	Gln	Val Phe Arg	Ile Ser	His Arg	Ser Tyr	Thr
	965		970			975
Phe Tyr His Glu	Asn	Pro Tyr Pro	Ser Ile	Phe Arg	Ile Asn	Phe
	980		985			990
Asp His Tyr Tyr	Pro	Val Pro Tyr	Ile Gln	Tyr Asp	Pro Leu	Leu
	995		1000			1005
Tyr Leu Arg Arg	Thr	Ser Asp Leu	Lys Ser	Lys Lys	Lys Arg	Gly
	1010		1015			1020
Arg Pro Ala Lys	Thr	Asn Asp Thr	Met Thr	Lys Val	Pro Phe	Leu
	1025		1030			1035
Gln Gly Phe Ser	Tyr	Pro Ile Pro	Ser Gly	Ser Tyr	Tyr Ala	Pro
	1040		1045			1050
Tyr Gly Met Pro	Tyr	Thr Ser Met	Pro Met	Met Asn	Leu Gly	Tyr
	1055		1060			1065
Tyr Gly Gln Tyr	Pro	Ala Pro Leu	Tyr Leu	Ser His	Thr Leu	Gly
	1070		1075			1080
Ala Ala Ser Pro	Phe	Met Arg Pro	Thr Val	Pro Pro	Pro Gln	Phe
	1085		1090			1095
His Thr Asn Ser	His	Val Lys Met	Ser Gly	Ala Ala	Lys His	Lys
	1100		1105			1110
Ala Lys His Gly	Val	His Leu Gln	Gly Pro	Val Ser	Met Gly	Leu
	1115		1120			1125
Gly Asp Met Gln	Pro	Ser Leu Asn	Pro Pro	Lys Val	Gly Ser	Ala
	1130		1135			1140
Ser Leu Ser Ser	Gly	Arg Leu His	Lys Arg	Lys His	Lys His	Lys
	1145		1150			1155
His Lys His Lys	Glu	Asp Arg Ile	Leu Gly	Thr His	Asp Asn	Leu
	1160		1165			1170
Ser Gly Leu Phe	Ala	Gly Lys Ala	Thr Gly	Phe Ser	Ser His	Ile
	1175		1180			1185
Leu Ser Glu Arg	Leu	Ser Ser Ala	Asp Lys	Glu Leu	Pro Leu	Val
	1190		1195			1200
Ser Glu Lys Asn	Lys	His Lys Glu	Lys Gln	Lys His	Gln His	Ser
	1205		1210			1215
Glu Ala Gly His	Lys	Ala Ser Lys	Asn Asn	Phe Glu	Val Asp	Thr

1220 1225 1230  
 Leu Ser Thr Leu Ser Leu Ser Asp Ala Gln His Trp Thr Gln Ala  
 1235 1240 1245  
 Lys Glu Lys Gly Asp Leu Ser Ser Glu Pro Val Asp Ser Cys Thr  
 1250 1255 1260  
 Lys Arg Tyr Ser Gly Ser Gly Gly Asp Gly Gly Ser Thr Arg Ser  
 1265 1270 1275  
 Glu Asn Leu Asp Val Phe Ser Glu Met Asn Pro Ser Asn Asp Lys  
 1280 1285 1290  
 Trp Asp Ser Asp Val Ser Gly Ser Lys Arg Arg Ser Tyr Glu Gly  
 1295 1300 1305  
 Phe Gly Thr Tyr Arg Glu Lys Asp Ile Gln Ala Phe Lys Met Asn  
 1310 1315 1320  
 Arg Lys Glu Arg Ser Ser Tyr Asp Ser Ser Met Ser Pro Gly Met  
 1325 1330 1335  
 Pro Ser Pro His Leu Lys Val Asp Gln Thr Ala Val His Ser Lys  
 1340 1345 1350  
 Asn Glu Gly Ser Val Pro Thr Met Met Thr Arg Lys Lys Pro Ala  
 1355 1360 1365  
 Ala Val Asp Ser Val Thr Ile Pro Pro Ala Pro Val Leu Ser Leu  
 1370 1375 1380  
 Leu Ala Ala Ser Ala Ala Thr Ser Asp Ala Val Gly Ser Ser Leu  
 1385 1390 1395  
 Lys Lys Arg Phe Lys Arg Arg Glu Ile Glu Ala Ile Gln Cys Glu  
 1400 1405 1410  
 Val Arg Lys Met Cys Asn Tyr Thr Lys Ile Leu Ser Thr Lys Lys  
 1415 1420 1425  
 Asn Leu Asp His Val Asn Lys Ile Leu Lys Ala Lys Arg Leu Gln  
 1430 1435 1440  
 Arg Gln Ser Lys Thr Gly Asn Asn Phe Val Lys Lys Arg Arg Gly  
 1445 1450 1455  
 Arg Pro Arg Lys Gln Pro Thr Gln Phe Asp Glu Asp Ser Arg Asp  
 1460 1465 1470  
 Gln Met Pro Val Leu Glu Lys Cys Ile Asp Leu Pro Ser Lys Arg  
 1475 1480 1485  
 Gly Gln Lys Pro Ser Leu Ser Pro Leu Val Leu Glu Pro Ala Ala  
 1490 1495 1500  
 Ser Gln Asp Thr Ile Met Ala Thr Ile Glu Ala Val Ile His Met  
 1505 1510 1515  
 Ala Arg Glu Ala Pro Pro Leu Pro Pro Pro Pro Pro Pro Leu  
 1520 1525 1530  
 Pro Pro Pro Pro Pro Pro Pro Leu Pro Pro Pro Pro Pro Leu Pro  
 1535 1540 1545  
 Lys Thr Pro Arg Gly Gly Lys Arg Lys His Lys Pro Gln Ala Pro  
 1550 1555 1560  
 Ala Gln Pro Pro Gln Gln Ser Pro Pro Gln Gln Pro Leu Pro Gln  
 1565 1570 1575  
 Glu Glu Glu Val Lys Ala Lys Arg Gln Arg Lys Ser Arg Gly Ser  
 1580 1585 1590  
 Glu Ser Glu Val Leu Pro  
 1595

<210> 19  
 <211> 191  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7506232CD1

<400> 19  
 Met Glu Ala Gln Ala Gln Gly Leu Leu Glu Thr Glu Pro Leu Gln

1	5	10	15
Gly Thr Asp Glu Asp	Ala Val Ala Ser	Ala Asp Phe Ser Ser	Met
20	25	30	
Leu Ser Glu Glu Glu	Lys Glu Glu Leu Lys	Ala Glu Leu Val Gln	
35	40	45	
Leu Glu Asp Glu Ile	Thr Thr Leu Arg Gln	Val Leu Ser Ala Lys	
50	55	60	
Glu Arg His Leu Val	Glu Ile Lys Gln Lys	Leu Gly Met Asn Leu	
65	70	75	
Met Asn Glu Leu Lys	Gln Asn Phe Ser Lys	Ser Trp His Asp Met	
80	85	90	
Gln Thr Thr Thr Ala	Tyr Lys Lys Thr His	Glu Thr Leu Ser His	
95	100	105	
Ala Gly Gln Lys Ala	Thr Ala Ala Phe Ser	Asn Val Gly Thr Ala	
110	115	120	
Ile Ser Lys Lys Phe	Gly Asp Met Arg Asn	Ser Pro Thr Phe Lys	
125	130	135	
Ser Phe Glu Glu Arg	Val Glu Thr Thr Val	Thr Ser Leu Lys Thr	
140	145	150	
Lys Val Gly Gly Thr	Asn Pro Asn Gly Gly	Ser Phe Glu Glu Val	
155	160	165	
Leu Ser Ser Thr Ala	His Ala Ser Ala Gln	Ser Leu Ala Gly Gly	
170	175	180	
Ser Arg Arg Thr Lys	Glu Glu Glu Leu Gln	Cys	
185	190		

&lt;210&gt; 20

&lt;211&gt; 143

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2585358CD1

&lt;400&gt; 20

Met Arg Asp Ala Asp	Ala Asp Ala Gly Gly	Gly Ala Asp Gly Gly
1	5	10
15		
Asp Gly Arg Gly Gly	His Ser Cys Arg Gly	Gly Val Asp Thr Ala
20	25	30
Ala Ala Pro Ala Gly	Gly Ala Pro Pro Ala	His Ala Pro Gly Pro
35	40	45
Gly Arg Asp Ala Ala	Ser Ala Ala Arg Gly	Ser Arg Met Arg Pro
50	55	60
His Ile Phe Thr Leu	Ser Val Pro Phe Pro	Thr Pro Leu Glu Ala
65	70	75
Glu Ile Ala His Gly	Ser Leu Ala Pro Asp	Ala Glu Pro His Gln
80	85	90
Arg Val Val Gly Lys	Asp Leu Thr Val Ser	Gly Arg Ile Leu Val
95	100	105
Val Arg Trp Lys Ala	Glu Asp Cys Arg Leu	Leu Arg Ile Ser Val
110	115	120
Ile Asn Phe Leu Asp	Gln Leu Ser Leu Val	Val Arg Thr Met Gln
125	130	135
Arg Phe Gly Pro Pro	Val Ser Arg	
140		

&lt;210&gt; 21

&lt;211&gt; 1488

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3961495CD1

&lt;400&gt; 21

```

Met Gln Thr Asn Gly Leu Pro Asn Trp Gly Met Ala Val Gly Met
  1          5          10          15
Gly Ala Ile Ile Pro Pro His Leu Gln Gly Leu Pro Gly Ala Asn
          20          25          30
Gly Ser Ser Val Ser Gln Val Ser Gly Gly Ser Ala Glu Gly Ile
          35          40          45
Ser Asn Ser Val Trp Gly Leu Ser Pro Gly Asn Pro Ala Thr Gly
          50          55          60
Asn Ser Asn Ser Gly Phe Ser Gln Gly Asn Gly Asp Thr Val Asn
          65          70          75
Ser Ala Leu Ser Ala Lys Gln Asn Gly Ser Ser Ser Ala Val Gln
          80          85          90
Lys Glu Gly Ser Gly Gly Asn Ala Trp Asp Ser Gly Pro Pro Ala
          95          100          105
Gly Pro Gly Ile Leu Ala Trp Gly Arg Gly Ser Gly Asn Asn Gly
          110          115          120
Val Gly Asn Ile His Ser Gly Ala Trp Gly His Pro Ser Arg Ser
          125          130          135
Thr Ser Asn Gly Val Asn Gly Glu Trp Gly Lys Pro Pro Asn Gln
          140          145          150
His Ser Asn Ser Asp Ile Asn Gly Lys Gly Ser Thr Gly Trp Glu
          155          160          165
Ser Pro Ser Val Thr Ser Gln Asn Pro Thr Val Gln Pro Gly Gly
          170          175          180
Glu His Met Asn Ser Trp Ala Lys Ala Ala Ser Ser Gly Thr Thr
          185          190          195
Ala Ser Glu Gly Ser Ser Asp Gly Ser Gly Asn His Asn Glu Gly
          200          205          210
Ser Thr Gly Arg Glu Gly Thr Gly Glu Gly Arg Arg Arg Asp Lys
          215          220          225
Gly Ile Ile Asp Gln Gly His Ile Gln Leu Pro Arg Asn Asp Leu
          230          235          240
Asp Pro Arg Val Leu Ser Asn Thr Gly Trp Gly Gln Thr Pro Val
          245          250          255
Lys Gln Asn Thr Ala Trp Glu Phe Glu Glu Ser Pro Arg Ser Glu
          260          265          270
Arg Lys Asn Asp Asn Gly Thr Glu Ala Trp Gly Cys Ala Ala Thr
          275          280          285
Gln Ala Ser Asn Ser Gly Gly Lys Asn Asp Gly Ser Ile Met Asn
          290          295          300
Ser Thr Asn Thr Ser Ser Val Ser Gly Trp Val Asn Ala Pro Pro
          305          310          315
Ala Ala Val Pro Ala Asn Thr Gly Trp Gly Asp Ser Asn Asn Lys
          320          325          330
Ala Pro Ser Gly Pro Gly Val Trp Gly Asp Ser Ile Ser Ser Thr
          335          340          345
Ala Val Ser Thr Ala Ala Ala Ala Lys Ser Gly His Ala Trp Ser
          350          355          360
Gly Ala Ala Asn Gln Glu Asp Lys Ser Pro Thr Trp Gly Glu Pro
          365          370          375
Pro Lys Pro Lys Ser Gln His Trp Gly Asp Gly Gln Arg Ser Asn
          380          385          390
Pro Ala Trp Ser Ala Gly Gly Gly Asp Trp Ala Asp Ser Ser Ser
          395          400          405
Val Leu Gly His Leu Gly Asp Gly Lys Lys Asn Gly Ser Gly Trp
          410          415          420
Asp Ala Asp Ser Asn Arg Ser Gly Ser Gly Trp Asn Asp Thr Thr
          425          430          435
Arg Ser Gly Asn Ser Gly Trp Gly Asn Ser Thr Asn Thr Lys Ala

```



Asn Pro Gly Thr	440	445	450
Asn Trp Gly Glu Thr	455	460	465
Gln Asn Trp Ala	470	475	480
Gly Gly Ala Ala	485	490	495
Gly Pro Val Pro	500	505	510
Trp Glu Glu Pro	515	520	525
Asp Asp Gly Thr	530	535	540
Lys Thr Val Asn	545	550	555
Ser Thr Thr Thr	560	565	570
Asn Thr Thr His	575	580	585
Thr Gln Leu Asn	590	595	600
Trp Gly Glu Met	605	610	615
Gly Glu Pro Ser	620	625	630
Ala Trp Gly Lys	635	640	645
Pro Ala Glu Pro	650	655	660
Ala Ala Ser Ala	665	670	675
Gly Trp Gly Ser	680	685	690
Trp Glu Asp Glu	695	700	705
Glu Ser Thr Ser	710	715	720
Gly Leu Gln Lys	725	730	735
Trp Ile Met Ser	740	745	750
Pro Arg Glu Pro	755	760	765
Leu Asp Gln Ala	770	775	780
Asp Lys Arg Gly	785	790	795
Lys Pro Leu Gly	800	805	810
Asp Arg Pro Thr	815	820	825
Pro Thr Pro Ser	830	835	840
Leu Ser His Ser	845	850	855
Ser Gly Leu Gly	860	865	870
Ser Gly Asn Leu	875	880	885
Arg Thr Met Gln	890	895	900
Ser Ser Gln Pro	905	910	915

Pro Gln Val Gln Ala Gln Leu Leu Gln Phe Ala Ala Lys Asn Ile  
 920 925 930  
 Gly Leu Asn Pro Ala Leu Leu Thr Ser Pro Ile Asn Pro Gln His  
 935 940 945  
 Met Thr Met Leu Asn Gln Leu Tyr Gln Leu Gln Leu Ala Tyr Gln  
 950 955 960  
 Arg Leu Gln Ile Gln Gln Gln Met Leu Gln Ala Gln Arg Asn Val  
 965 970 975  
 Ser Gly Ser Met Arg Gln Gln Glu Gln Gln Val Ala Arg Thr Ile  
 980 985 990  
 Thr Asn Leu Gln Gln Gln Ile Gln Gln His Gln Arg Gln Leu Ala  
 995 1000 1005  
 Gln Ala Leu Leu Val Lys Gln Pro Pro Pro Pro Pro Pro Pro  
 1010 1015 1020  
 His Leu Ser Leu His Pro Ser Ala Gly Lys Ser Ala Met Asp Ser  
 1025 1030 1035  
 Phe Pro Ser His Pro Gln Thr Pro Gly Leu Pro Asp Leu Gln Thr  
 1040 1045 1050  
 Lys Glu Gln Gln Ser Ser Pro Asn Thr Phe Ala Pro Tyr Pro Leu  
 1055 1060 1065  
 Ala Gly Leu Asn Pro Asn Met Asn Val Asn Ser Met Asp Met Thr  
 1070 1075 1080  
 Gly Gly Leu Ser Val Lys Asp Pro Ser Gln Ser Gln Ser Arg Leu  
 1085 1090 1095  
 Pro Gln Trp Thr His Pro Asn Ser Met Asp Asn Leu Pro Ser Ala  
 1100 1105 1110  
 Ala Ser Pro Leu Glu Gln Asn Pro Ser Lys His Gly Ala Ile Pro  
 1115 1120 1125  
 Gly Gly Leu Ser Ile Gly Pro Pro Gly Lys Ser Ser Ile Asp Asp  
 1130 1135 1140  
 Ser Tyr Gly Arg Tyr Asp Leu Ile Gln Asn Ser Glu Ser Pro Ala  
 1145 1150 1155  
 Ser Pro Pro Val Ala Val Pro His Ser Arg Ser Arg Ala Lys Ser  
 1160 1165 1170  
 Asp Ser Asp Lys Ile Ser Asn Gly Pro Ser Ile Asn Trp Pro Pro  
 1175 1180 1185  
 Glu Phe His Pro Gly Val Pro Trp Lys Gly Leu Gln Asn Ile Asp  
 1190 1195 1200  
 Pro Glu Asn Asp Pro Asp Val Thr Pro Gly Thr Val Pro Thr Gly  
 1205 1210 1215  
 Pro Thr Ile Asn Thr Thr Ile Gln Asp Val Asn Arg Tyr Leu Leu  
 1220 1225 1230  
 Lys Ser Gly Gly Lys Leu Ser Asp Ile Lys Ser Thr Trp Ser Ser  
 1235 1240 1245  
 Gly Pro Thr Ser His Thr Gln Ala Ser Leu Ser His Glu Leu Trp  
 1250 1255 1260  
 Lys Val Pro Arg Asn Ser Thr Ala Pro Thr Arg Pro Pro Pro Gly  
 1265 1270 1275  
 Leu Thr Asn Pro Lys Pro Ser Ser Thr Trp Gly Ala Ser Pro Leu  
 1280 1285 1290  
 Gly Trp Thr Ser Ser Tyr Ser Ser Gly Ser Ala Trp Ser Thr Asp  
 1295 1300 1305  
 Thr Ser Gly Arg Thr Ser Ser Trp Leu Val Leu Arg Asn Leu Thr  
 1310 1315 1320  
 Pro Gln Ile Asp Gly Ser Thr Leu Arg Thr Leu Cys Leu Gln His  
 1325 1330 1335  
 Gly Pro Leu Ile Thr Phe His Leu Asn Leu Thr Gln Gly Asn Ala  
 1340 1345 1350  
 Val Val Arg Tyr Ser Ser Lys Glu Glu Ala Ala Lys Ala Gln Lys  
 1355 1360 1365  
 Ser Leu His Met Cys Val Leu Gly Asn Thr Thr Ile Leu Ala Glu  
 1370 1375 1380  
 Phe Ala Gly Glu Glu Glu Val Asn Arg Phe Leu Ala Gln Gly Gln

1385	1390	1395
Ala Leu Pro Pro Thr Ser Ser Trp Gln Ser Ser Ser Ala Ser Ser		
1400	1405	1410
Gln Pro Arg Leu Ser Ala Ala Gly Ser Ser His Gly Leu Val Arg		
1415	1420	1425
Ser Asp Ala Gly His Trp Asn Ala Pro Cys Leu Gly Gly Lys Gly		
1430	1435	1440
Ser Ser Glu Leu Leu Trp Gly Gly Val Pro Gln Tyr Ser Ser Ser		
1445	1450	1455
Leu Trp Gly Pro Pro Ser Ala Asp Asp Ser Arg Val Ile Gly Ser		
1460	1465	1470
Pro Thr Pro Leu Thr Thr Leu Leu Pro Gly Asp Leu Leu Ser Gly		
1475	1480	1485
Glu Ser Leu		

<210> 22  
 <211> 505  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7500801CD1

<400> 22

Met Gly Gly Ala Gly Ile Leu Leu Leu Leu Leu Ala Gly Ala Gly	1	5	10	15
Val Val Val Ala Trp Arg Pro Pro Lys Gly Lys Cys Pro Leu Arg	20	25	30	
Cys Ser Cys Ser Lys Asp Ser Ala Leu Cys Glu Gly Ser Pro Asp	35	40	45	
Leu Pro Val Ser Phe Ser Pro Thr Leu Leu Ser Leu Ser Leu Val	50	55	60	
Arg Thr Gly Val Thr Gln Leu Lys Ala Gly Ser Phe Leu Arg Ile	65	70	75	
Pro Ser Leu His Leu Leu Leu Phe Thr Ser Asn Ser Phe Ser Val	80	85	90	
Ile Glu Asp Asp Ala Phe Ala Gly Leu Ser His Leu Gln Tyr Leu	95	100	105	
Phe Ile Glu Asp Asn Glu Ile Gly Ser Ile Ser Lys Asn Ala Leu	110	115	120	
Arg Gly Leu Arg Ser Leu Thr His Leu Ser Leu Ala Asn Asn His	125	130	135	
Leu Glu Thr Leu Pro Arg Phe Leu Phe Arg Gly Leu Asp Thr Leu	140	145	150	
Thr His Val Asp Leu Arg Gly Asn Pro Phe Gln Cys Asp Cys Arg	155	160	165	
Val Leu Trp Leu Leu Gln Trp Met Pro Thr Val Asn Ala Ser Val	170	175	180	
Gly Thr Gly Ala Cys Ala Gly Pro Ala Ser Leu Ser His Met Gln	185	190	195	
Leu His His Leu Asp Pro Lys Thr Phe Lys Cys Arg Ala Ile Glu	200	205	210	
Leu Ser Trp Phe Gln Thr Val Gly Glu Ser Ala Leu Ser Val Glu	215	220	225	
Pro Phe Ser Tyr Gln Gly Glu Pro His Ile Val Leu Ala Gln Pro	230	235	240	
Phe Ala Gly Arg Cys Leu Ile Leu Ser Trp Asp Tyr Ser Leu Gln	245	250	255	
Arg Phe Arg Pro Glu Glu Glu Leu Pro Ala Ala Ser Val Val Ser	260	265	270	
Cys Lys Pro Leu Val Leu Gly Pro Ser Leu Phe Val Leu Ala Ala				

275	280	285
Arg Leu Trp Gly	Gly Ser Gln Leu Trp	Ala Arg Pro Ser Pro Gly
290	295	300
Leu Arg Leu Ala	Pro Thr Gln Thr Leu	Ala Pro Arg Arg Leu Leu
305	310	315
Arg Pro Asn Asp	Ala Glu Leu Leu Trp	Leu Glu Gly Gln Pro Cys
320	325	330
Phe Val Val Ala	Asp Ala Ser Lys Ala	Gly Ser Thr Thr Leu Leu
335	340	345
Cys Arg Asp Gly	Pro Gly Phe Tyr Pro	His Gln Ser Leu His Ala
350	355	360
Trp His Arg Asp	Thr Asp Ala Glu Ala	Leu Glu Leu Asp Gly Arg
365	370	375
Pro His Leu Leu	Leu Ala Ser Ala Ser	Gln Arg Pro Val Leu Phe
380	385	390
His Trp Thr Gly	Gly Arg Phe Glu Arg	Arg Thr Asp Ile Pro Glu
395	400	405
Ala Glu Asp Val	Tyr Ala Thr Arg His	Phe Gln Ala Gly Gly Asp
410	415	420
Val Phe Leu Cys	Leu Thr Arg Tyr Ile	Gly Asp Ser Met Val Met
425	430	435
Arg Trp Asp Gly	Ser Met Phe Arg Leu	Leu Gln Gln Leu Pro Ser
440	445	450
Arg Gly Ala His	Gly Pro Ser Leu Leu	Asp Leu Pro Ser Cys Pro
455	460	465
Ser Gly Leu Lys	Gln His Leu Asp Ile	Ser Ala Gly Ala Asp Ser
470	475	480
Pro Ser Lys Ala	Thr Trp Gly Val Gln	Ala Gly Ala Leu His Gly
485	490	495
Val Glu Val Glu	Val Thr Cys Arg Ala	Pro
500	505	

&lt;210&gt; 23

&lt;211&gt; 1544

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506414CD1

&lt;400&gt; 23

Met Leu Lys Cys Ile	Pro Leu Trp Arg Cys	Asn Arg His Val Glu
1	5	10
Ser Val Asp Lys Arg	His Cys Ser Leu Gln	Val Val Pro Glu Glu
20	25	30
Ile Tyr Arg Tyr Ser	Arg Ser Leu Glu Glu	Leu Leu Leu Asp Ala
35	40	45
Asn Gln Leu Arg Glu	Leu Pro Lys Pro Phe	Phe Arg Leu Leu Asn
50	55	60
Leu Arg Lys Leu Gly	Leu Ser Asp Asn Glu	Ile Gln Arg Leu Pro
65	70	75
Pro Glu Val Ala Asn	Phe Met Gln Leu Val	Glu Leu Asp Val Ser
80	85	90
Arg Asn Asp Ile Pro	Glu Ile Pro Glu Ser	Ile Lys Phe Cys Lys
95	100	105
Ala Leu Glu Ile Ala	Asp Phe Ser Gly Asn	Pro Leu Ser Arg Leu
110	115	120
Pro Asp Gly Phe Thr	Gln Leu Arg Ser Leu	Ala His Leu Ala Leu
125	130	135
Asn Asp Val Ser Leu	Gln Ala Leu Pro Gly	Asp Val Gly Asn Leu
140	145	150
Ala Asn Leu Val Thr	Leu Glu Leu Arg Glu	Asn Leu Leu Lys Ser

Leu Pro Ala Ser	155	160	165
Leu Ser Phe Leu Val	170	Lys Leu Glu Gln Leu Asp	180
Leu Gly Gly Asn	185	175	185
Asp Leu Glu Val Leu	190	Pro Asp Thr Leu Gly Ala	195
Leu Pro Asn Leu	200	190	200
Arg Glu Leu Trp Leu	205	Asp Arg Asn Gln Leu Ser	210
Ala Leu Pro Pro	215	205	215
Glu Leu Gly Asn Leu	220	Arg Arg Leu Val Cys Leu	225
Asp Val Ser Glu	230	220	230
Asn Arg Leu Glu Glu	235	Leu Pro Ala Glu Leu Gly	240
Gly Leu Val Leu	245	235	245
Leu Thr Asp Leu Leu	250	Leu Ser Gln Asn Leu Leu	255
Arg Arg Leu Pro	260	250	260
Asp Gly Ile Gly Gln	265	Leu Lys Gln Leu Ser Ile	270
Leu Lys Val Asp	275	265	275
Gln Asn Arg Leu Cys	280	Glu Val Thr Glu Ala Ile	285
Gly Asp Cys Glu	290	280	290
Asn Leu Ser Glu Leu	295	Ile Leu Thr Glu Asn Leu	300
Leu Met Ala Leu	305	295	305
Pro Arg Ser Leu Gly	310	Lys Leu Thr Lys Leu Thr	315
Asn Leu Asn Val	320	310	320
Asp Arg Asn His Leu	325	Glu Ala Leu Pro Pro Glu	330
Ile Gly Gly Cys	335	325	335
Val Ala Leu Ser Val	340	Leu Ser Leu Arg Asp Asn	345
Arg Leu Ala Val	350	340	350
Leu Pro Pro Glu Leu	355	Ala His Thr Thr Glu Leu	360
His Val Leu Asp	365	355	365
Val Ala Gly Asn Arg	370	Leu Gln Ser Leu Pro Phe	375
Ala Leu Thr His	380	370	380
Leu Asn Leu Lys Ala	385	Leu Trp Leu Ala Glu Asn	390
Gln Ala Gln Pro	395	385	395
Met Leu Arg Phe Gln	400	Thr Glu Asp Asp Ala Arg	405
Thr Gly Glu Lys	410	400	410
Val Leu Thr Cys Tyr	415	Leu Leu Pro Gln Gln Pro	420
Pro Leu Ser Leu	425	415	425
Glu Asp Ala Gly Gln	430	Gln Gly Ser Leu Ser Glu	435
Thr Trp Ser Asp	440	430	440
Ala Pro Pro Ser Arg	445	Val Ser Val Ile Gln Phe	450
Leu Glu Ala Pro	455	445	455
Ile Gly Asp Glu Asp	460	Ala Glu Glu Ala Ala Ala	465
Glu Lys Arg Gly	470	460	470
Leu Gln Arg Arg Ala	475	Thr Pro His Pro Ser Glu	480
Leu Lys Val Met	485	475	485
Lys Arg Ser Ile Glu	490	Gly Arg Arg Ser Glu Ala	495
Cys Pro Cys Gln	500	490	500
Pro Asp Ser Gly Ser	505	Pro Leu Pro Ala Glu Glu	510
Glu Lys Arg Leu	515	505	515
Ser Ala Glu Ser Gly	520	Leu Ser Glu Asp Ser Arg	525
Pro Ser Ala Ser	530	520	530
Thr Val Ser Glu Ala	535	Glu Pro Glu Gly Pro Ser	540
Ala Glu Ala Gln	545	535	545
Gly Gly Ser Gln Gln	550	Glu Ala Thr Thr Ala Gly	555
Gly Glu Glu Asp	560	550	560
Ala Glu Glu Asp Tyr	565	Gln Glu Pro Thr Val His	570
Phe Ala Glu Asp	575	565	575
Ala Leu Leu Pro Gly	580	Asp Asp Arg Glu Ile Glu	585
Glu Gly Gln Pro	590	580	590
Glu Ala Pro Trp Thr	595	Leu Pro Gly Gly Arg Gln	600
Arg Leu Ile Arg	605	595	605
Lys Asp Thr Pro His	610	Tyr Lys Lys His Phe Lys	615
Ile Ser Lys Leu	620	610	620
Pro Gln Pro Glu Ala	625	Val Val Ala Leu Leu Gln	630

Gly Met Gln Pro Asp	Gly Glu Gly Pro Val	Ala Pro Gly Gly Trp
635	640	645
His Asn Gly Pro His	Ala Pro Trp Ala Pro	Arg Ala Gln Lys Glu
650	655	660
Glu Glu Glu Glu Glu	Glu Gly Ser Pro Gln	Glu Glu Glu Glu Glu
665	670	675
Glu Glu Glu Glu Asn	Arg Ala Glu Glu Glu	Ala Ser Thr Glu
680	685	690
Glu Glu Asp Lys Glu	Gly Ala Val Val Ser	Ala Pro Ser Val Lys
695	700	705
Gly Val Ser Phe Asp	Gln Ala Asn Asn Leu	Leu Ile Glu Pro Ala
710	715	720
Arg Ile Glu Glu Glu	Glu Leu Thr Leu Thr	Ile Leu Arg Gln Thr
725	730	735
Gly Gly Leu Gly Ile	Ser Ile Ala Gly Gly	Lys Gly Ser Thr Pro
740	745	750
Tyr Lys Gly Asp Asp	Glu Gly Ile Phe Ile	Ser Arg Val Ser Glu
755	760	765
Glu Gly Pro Ala Ala	Arg Ala Gly Val Arg	Val Gly Asp Lys Leu
770	775	780
Leu Glu Val Asn Gly	Val Ala Leu Gln Gly	Ala Glu His His Glu
785	790	795
Ala Val Glu Ala Leu	Arg Gly Ala Gly Thr	Ala Val Gln Met Arg
800	805	810
Val Trp Arg Glu Arg	Met Val Glu Pro Glu	Asn Ala Val Thr Ile
815	820	825
Thr Pro Leu Arg Pro	Glu Asp Asp Tyr Ser	Pro Arg Glu Arg Arg
830	835	840
Gly Gly Gly Leu Arg	Leu Pro Leu Leu Pro	Pro Glu Ser Pro Gly
845	850	855
Pro Leu Arg Gln Arg	His Val Ala Cys Leu	Ala Arg Ser Glu Arg
860	865	870
Gly Leu Gly Phe Ser	Ile Ala Gly Gly Lys	Gly Ser Thr Pro Tyr
875	880	885
Arg Ala Gly Asp Ala	Gly Ile Phe Val Ser	Arg Ile Ala Glu Gly
890	895	900
Gly Ala Ala His Arg	Ala Gly Thr Leu Gln	Val Gly Asp Arg Val
905	910	915
Leu Ser Glu Ile Arg	Leu Pro Arg Ala Gly	Gly Pro Leu Gly Leu
920	925	930
Ser Ile Val Gly Gly	Ser Asp His Ser Ser	His Pro Phe Gly Val
935	940	945
Gln Glu Pro Gly Val	Phe Ile Ser Lys Val	Leu Pro Arg Gly Leu
950	955	960
Ala Ala Arg Ser Gly	Leu Arg Val Gly Asp	Arg Ile Leu Ala Val
965	970	975
Asn Gly Gln Asp Val	Arg Asp Ala Thr His	Gln Glu Ala Val Ser
980	985	990
Ala Leu Leu Arg Pro	Cys Leu Glu Leu Ser	Leu Leu Val Arg Arg
995	1000	1005
Asp Pro Ala Pro Pro	Gly Leu Arg Glu Leu	Cys Ile Gln Lys Ala
1010	1015	1020
Pro Gly Glu Arg Leu	Gly Ile Ser Ile Arg	Gly Gly Ala Arg Gly
1025	1030	1035
His Ala Gly Asn Pro	Arg Asp Pro Thr Asp	Glu Gly Ile Phe Ile
1040	1045	1050
Ser Lys Val Ser Pro	Thr Gly Ala Ala Gly	Arg Asp Gly Arg Leu
1055	1060	1065
Arg Val Gly Leu Arg	Leu Leu Glu Val Asn	Gln Gln Ser Leu Leu
1070	1075	1080
Gly Leu Thr His Gly	Glu Ala Val Gln Leu	Leu Arg Ser Val Gly
1085	1090	1095
Asp Thr Leu Thr Val	Leu Val Cys Asp Gly	Phe Glu Ala Ser Thr

1100	1105	1110
Asp Ala Ala Leu Glu Val Ser Pro Gly Val Ile Ala Asn Pro Phe		
1115	1120	1125
Ala Ala Gly Ile Gly His Arg Asn Ser Leu Glu Ser Ile Ser Ser		
1130	1135	1140
Ile Asp Arg Glu Leu Ser Pro Glu Gly Pro Gly Lys Glu Lys Glu		
1145	1150	1155
Leu Pro Gly Gln Thr Leu His Trp Gly Pro Glu Ala Thr Glu Ala		
1160	1165	1170
Ala Gly Arg Gly Leu Gln Pro Leu Lys Leu Asp Tyr Arg Ala Leu		
1175	1180	1185
Ala Ala Val Pro Ser Ala Gly Ser Val Gln Arg Val Pro Ser Gly		
1190	1195	1200
Ala Ala Gly Gly Lys Met Ala Glu Ser Pro Cys Ser Pro Ser Gly		
1205	1210	1215
Gln Gln Pro Pro Ser Pro Pro Ser Pro Asp Glu Leu Pro Ala Asn		
1220	1225	1230
Val Lys Gln Ala Tyr Arg Ala Phe Ala Ala Val Pro Thr Ser His		
1235	1240	1245
Pro Pro Glu Asp Ala Pro Ala Gln Pro Pro Thr Pro Gly Pro Ala		
1250	1255	1260
Ala Ser Pro Glu Gln Leu Ser Phe Arg Glu Arg Gln Lys Tyr Phe		
1265	1270	1275
Glu Leu Glu Val Arg Val Pro Gln Ala Glu Gly Pro Pro Lys Arg		
1280	1285	1290
Val Ser Leu Val Gly Ala Asp Asp Leu Arg Lys Met Gln Glu Glu		
1295	1300	1305
Glu Ala Arg Lys Leu Gln Gln Lys Arg Ala Gln Met Leu Arg Glu		
1310	1315	1320
Ala Ala Glu Ala Gly Ala Glu Ala Arg Leu Ala Leu Asp Gly Glu		
1325	1330	1335
Thr Leu Gly Glu Glu Glu Gln Glu Asp Glu Gln Pro Pro Trp Ala		
1340	1345	1350
Ser Pro Ser Pro Thr Ser Arg Gln Ser Pro Ala Ser Pro Pro Pro		
1355	1360	1365
Leu Gly Gly Gly Ala Pro Val Arg Thr Ala Lys Ala Glu Arg Arg		
1370	1375	1380
His Gln Glu Arg Leu Arg Val Gln Ser Pro Glu Pro Pro Ala Pro		
1385	1390	1395
Glu Arg Ala Leu Ser Pro Ala Glu Leu Arg Ala Leu Glu Ala Glu		
1400	1405	1410
Lys Arg Ala Leu Trp Arg Ala Ala Arg Met Lys Ser Leu Glu Gln		
1415	1420	1425
Asp Ala Leu Arg Ala Gln Met Val Leu Ser Arg Ser Gln Glu Gly		
1430	1435	1440
Arg Gly Thr Arg Gly Pro Leu Glu Arg Leu Ala Glu Ala Pro Ser		
1445	1450	1455
Pro Ala Pro Thr Pro Ser Pro Thr Pro Val Glu Asp Leu Gly Pro		
1460	1465	1470
Gln Thr Ser Thr Ser Pro Gly Arg Leu Ser Pro Asp Phe Ala Glu		
1475	1480	1485
Glu Leu Arg Ser Leu Glu Pro Ser Pro Ser Pro Gly Pro Gln Glu		
1490	1495	1500
Glu Asp Gly Glu Val Ala Leu Val Leu Leu Gly Arg Pro Ser Pro		
1505	1510	1515
Gly Ala Val Gly Pro Glu Asp Val Ala Leu Cys Ser Ser Arg Arg		
1520	1525	1530
Pro Val Arg Pro Gly Arg Arg Gly Leu Gly Pro Val Pro Ser		
1535	1540	

<210> 24  
 <211> 1609  
 <212> PRT

$\langle 220 \rangle$ 

<223> Incyte ID No: 7506415CD1

<400> 24

34/89



Pro Leu Ser Leu Glu Asp Ala Gly Gln Gln Gly Ser Leu Ser Glu  
 425 430 435  
 Thr Trp Ser Asp Ala Pro Pro Ser Arg Val Ser Val Ile Gln Phe  
 440 445 450  
 Leu Glu Ala Pro Ile Gly Asp Glu Asp Ala Glu Glu Ala Ala  
 455 460 465  
 Glu Lys Arg Gly Leu Gln Arg Arg Ala Thr Pro His Pro Ser Glu  
 470 475 480  
 Leu Lys Val Met Lys Arg Ser Ile Glu Gly Arg Arg Ser Glu Ala  
 485 490 495  
 Cys Pro Cys Gln Pro Asp Ser Gly Ser Pro Leu Pro Ala Glu Glu  
 500 505 510  
 Glu Lys Arg Leu Ser Ala Glu Ser Gly Leu Ser Glu Asp Ser Arg  
 515 520 525  
 Pro Ser Ala Ser Thr Val Ser Glu Ala Glu Pro Glu Gly Pro Ser  
 530 535 540  
 Ala Glu Ala Gln Gly Gly Ser Gln Gln Glu Ala Thr Thr Ala Gly  
 545 550 555  
 Gly Glu Glu Asp Ala Glu Glu Asp Tyr Gln Glu Pro Thr Val His  
 560 565 570  
 Phe Ala Glu Asp Ala Leu Leu Pro Gly Asp Asp Arg Glu Ile Glu  
 575 580 585  
 Glu Gly Gln Pro Glu Ala Pro Trp Thr Leu Pro Gly Gly Arg Gln  
 590 595 600  
 Arg Leu Ile Arg Lys Asp Thr Pro His Tyr Lys Lys His Phe Lys  
 605 610 615  
 Ile Ser Lys Leu Pro Gln Pro Glu Ala Val Val Ala Leu Leu Gln  
 620 625 630  
 Gly Met Gln Pro Asp Gly Glu Gly Pro Val Ala Pro Gly Gly Trp  
 635 640 645  
 His Asn Gly Pro His Ala Pro Trp Ala Pro Arg Ala Gln Lys Glu  
 650 655 660  
 Glu Glu Glu Glu Glu Glu Gly Ser Pro Gln Glu Glu Glu Glu  
 665 670 675  
 Glu Glu Glu Glu Asn Arg Ala Glu Glu Glu Glu Ala Ser Thr Glu  
 680 685 690  
 Glu Glu Asp Lys Glu Gly Ala Val Val Ser Ala Pro Ser Val Lys  
 695 700 705  
 Leu Thr Leu Thr Ile Leu Arg Gln Thr Gly Gly Leu Gly Ile Ser  
 710 715 720  
 Ile Ala Gly Gly Lys Gly Ser Thr Pro Tyr Lys Gly Asp Asp Glu  
 725 730 735  
 Gly Ile Phe Ile Ser Arg Val Ser Glu Glu Gly Pro Ala Ala Arg  
 740 745 750  
 Ala Gly Val Arg Val Gly Asp Lys Leu Leu Glu Val Asn Gly Val  
 755 760 765  
 Ala Leu Gln Gly Ala Glu His His Glu Ala Val Glu Ala Leu Arg  
 770 775 780  
 Gly Ala Gly Thr Ala Val Gln Met Arg Val Trp Arg Glu Arg Met  
 785 790 795  
 Val Glu Pro Glu Asn Ala Val Thr Ile Thr Pro Leu Arg Pro Glu  
 800 805 810  
 Asp Asp Tyr Ser Pro Arg Glu Arg Arg Gly Gly Gly Leu Arg Leu  
 815 820 825  
 Pro Leu Leu Pro Pro Glu Ser Pro Gly Pro Leu Arg Gln Arg His  
 830 835 840  
 Val Ala Cys Leu Ala Arg Ser Glu Arg Gly Leu Gly Phe Ser Ile  
 845 850 855  
 Ala Gly Gly Lys Gly Ser Thr Pro Tyr Arg Ala Gly Asp Ala Gly  
 860 865 870  
 Ile Phe Val Ser Arg Ile Ala Glu Gly Gly Ala Ala His Arg Ala  
 875 880 885  
 Gly Thr Leu Gln Val Gly Asp Arg Val Leu Ser Ile Asn Gly Val

890	895	900
Asp Val Thr Glu Ala Arg His Asp His Ala Val Ser Leu Leu Thr		
905	910	915
Ala Ala Ser Pro Thr Ile Ala Leu Leu Leu Glu Arg Glu Ala Gly		
920	925	930
Gly Pro Leu Pro Pro Ser Pro Leu Pro His Ser Ser Pro Pro Thr		
935	940	945
Ala Ala Val Ala Thr Thr Ser Ile Thr Thr Ala Thr Pro Gly Val		
950	955	960
Pro Gly Leu Pro Ser Leu Ala Pro Ser Leu Leu Ala Ala Ala Leu		
965	970	975
Glu Gly Pro Tyr Pro Val Glu Glu Ile Arg Leu Pro Arg Ala Gly		
980	985	990
Gly Pro Leu Gly Leu Ser Ile Val Gly Gly Ser Asp His Ser Ser		
995	1000	1005
His Pro Phe Gly Val Gln Glu Pro Gly Val Phe Ile Ser Lys Val		
1010	1015	1020
Leu Pro Arg Gly Leu Ala Ala Arg Ser Gly Leu Arg Val Gly Asp		
1025	1030	1035
Arg Ile Leu Ala Val Asn Gly Gln Asp Val Arg Asp Ala Thr His		
1040	1045	1050
Gln Glu Ala Val Ser Ala Leu Leu Arg Pro Cys Leu Glu Leu Ser		
1055	1060	1065
Leu Leu Val Arg Arg Asp Pro Ala Pro Pro Gly Leu Arg Glu Leu		
1070	1075	1080
Cys Ile Gln Lys Ala Pro Gly Glu Arg Leu Gly Ile Ser Ile Arg		
1085	1090	1095
Gly Gly Ala Arg Gly His Ala Gly Asn Pro Arg Asp Pro Thr Asp		
1100	1105	1110
Glu Gly Ile Phe Ile Ser Lys Val Ser Pro Thr Gly Ala Ala Gly		
1115	1120	1125
Arg Asp Gly Arg Leu Arg Val Gly Leu Arg Leu Leu Glu Val Asn		
1130	1135	1140
Gln Gln Ser Leu Leu Gly Leu Thr His Gly Glu Ala Val Gln Leu		
1145	1150	1155
Leu Arg Ser Val Gly Asp Thr Leu Thr Val Leu Val Cys Asp Gly		
1160	1165	1170
Phe Glu Ala Ser Thr Asp Ala Ala Leu Glu Val Ser Pro Gly Val		
1175	1180	1185
Ile Ala Asn Pro Phe Ala Ala Gly Ile Gly His Arg Asn Ser Leu		
1190	1195	1200
Glu Ser Ile Ser Ser Ile Asp Arg Glu Leu Ser Pro Glu Gly Pro		
1205	1210	1215
Gly Lys Glu Lys Glu Leu Pro Gly Gln Thr Leu His Trp Gly Pro		
1220	1225	1230
Glu Ala Thr Glu Ala Ala Gly Arg Gly Leu Gln Pro Leu Lys Leu		
1235	1240	1245
Asp Tyr Arg Ala Leu Ala Ala Val Pro Ser Ala Gly Ser Val Gln		
1250	1255	1260
Arg Val Pro Ser Gly Ala Ala Gly Gly Lys Met Ala Glu Ser Pro		
1265	1270	1275
Cys Ser Pro Ser Gly Gln Gln Pro Pro Ser Pro Pro Ser Pro Asp		
1280	1285	1290
Glu Leu Pro Ala Asn Val Lys Gln Ala Tyr Arg Ala Phe Ala Ala		
1295	1300	1305
Val Pro Thr Ser His Pro Pro Glu Asp Ala Pro Ala Gln Pro Pro		
1310	1315	1320
Thr Pro Gly Pro Ala Ala Ser Pro Glu Gln Leu Ser Phe Arg Glu		
1325	1330	1335
Arg Gln Lys Tyr Phe Glu Leu Glu Val Arg Val Pro Gln Ala Glu		
1340	1345	1350
Gly Pro Pro Lys Arg Val Ser Leu Val Gly Ala Asp Asp Leu Arg		
1355	1360	1365

Lys Met Gln Glu Glu Glu Ala Arg Lys Leu Gln Gln Lys Arg Ala  
 1370 1375 1380  
 Gln Met Leu Arg Glu Ala Ala Glu Ala Gly Ala Glu Ala Arg Leu  
 1385 1390 1395  
 Ala Leu Asp Gly Glu Thr Leu Gly Glu Glu Gln Glu Asp Glu  
 1400 1405 1410  
 Gln Pro Pro Trp Ala Ser Pro Ser Pro Thr Ser Arg Gln Ser Pro  
 1415 1420 1425  
 Ala Ser Pro Pro Pro Leu Gly Gly Gly Ala Pro Val Arg Thr Ala  
 1430 1435 1440  
 Lys Ala Glu Arg Arg His Gln Glu Arg Leu Arg Val Gln Ser Pro  
 1445 1450 1455  
 Glu Pro Pro Ala Pro Glu Arg Ala Leu Ser Pro Ala Glu Leu Arg  
 1460 1465 1470  
 Ala Leu Glu Ala Glu Lys Arg Ala Leu Trp Arg Ala Ala Arg Met  
 1475 1480 1485  
 Lys Ser Leu Glu Gln Asp Ala Leu Arg Ala Gln Met Val Leu Ser  
 1490 1495 1500  
 Arg Ser Gln Glu Gly Arg Gly Thr Arg Gly Pro Leu Glu Arg Leu  
 1505 1510 1515  
 Ala Glu Ala Pro Ser Pro Ala Pro Thr Pro Ser Pro Thr Pro Val  
 1520 1525 1530  
 Glu Asp Leu Gly Pro Gln Thr Ser Thr Ser Pro Gly Arg Leu Ser  
 1535 1540 1545  
 Pro Asp Phe Ala Glu Glu Leu Arg Ser Leu Glu Pro Ser Pro Ser  
 1550 1555 1560  
 Pro Gly Pro Gln Glu Glu Asp Gly Glu Val Ala Leu Val Leu Leu  
 1565 1570 1575  
 Gly Arg Pro Ser Pro Gly Ala Val Gly Pro Glu Asp Val Ala Leu  
 1580 1585 1590  
 Cys Ser Ser Arg Arg Pro Val Arg Pro Gly Arg Arg Gly Leu Gly  
 1595 1600 1605  
 Pro Val Pro Ser

&lt;210&gt; 25

&lt;211&gt; 895

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 72192179CD1

&lt;400&gt; 25

Met Lys Lys Glu Val Leu Gln Ser Ser Arg Asp Ile Met Gln Ser  
 1 5 10 15  
 Lys Ser Ala Cys Glu Ile Lys Gln Ser His Gln Glu Cys Ser Thr  
 20 25 30  
 Gln Gln Thr Gln Gln Lys Lys Tyr Leu Glu Gln Leu His Leu Pro  
 35 40 45  
 Gln Ser Lys Pro Ile Ser Pro Asn Phe Lys Val Lys Thr Ile Lys  
 50 55 60  
 Leu Pro Thr Leu Asp His Thr Leu Asn Glu Thr Asp His Ser Tyr  
 65 70 75  
 Glu Ser His Lys Gln Gln Ser Glu Ile Asp Val Gln Thr Phe Thr  
 80 85 90  
 Lys Lys Gln Tyr Leu Lys Thr Lys Lys Thr Glu Ala Ser Thr Glu  
 95 100 105  
 Cys Ser His Lys Gln Ser Leu Ala Glu Arg His Tyr Gln Leu Pro  
 110 115 120  
 Lys Lys Glu Lys Arg Val Thr Val Gln Leu Pro Thr Glu Ser Ile  
 125 130 135

Gln Lys Asn Gln	Glu Asp Lys Leu Lys Met Val Pro Arg Lys Gln	140	145	150
Arg Glu Phe Ser	Gly Ser Asp Arg Gly Lys Leu Pro Gly Ser Glu	155	160	165
Glu Lys Asn Gln	Gly Pro Ser Met Ile Gly Arg Lys Glu Glu Arg	170	175	180
Leu Ile Thr Glu	Arg Lys His Glu His Leu Lys Asn Lys Ser Ala	185	190	195
Pro Lys Val Val	Lys Gln Lys Val Ile Asp Ala His Leu Asp Ser	200	205	210
Gln Thr Gln Asn	Phe Gln Gln Thr Gln Ile Gln Thr Ala Glu Ser	215	220	225
Lys Ala Glu His	Lys Lys Leu Pro Gln Pro Tyr Asn Ser Leu Gln	230	235	240
Glu Glu Lys Cys	Leu Glu Val Lys Gly Ile Gln Glu Lys Gln Val	245	250	255
Phe Ser Asn Thr	Lys Asp Ser Lys Gln Glu Ile Thr Gln Asn Lys	260	265	270
Ser Phe Phe Ser	Ser Val Lys Glu Ser Gln Arg Asp Asp Gly Lys	275	280	285
Gly Ala Leu Asn	Ile Val Glu Phe Leu Arg Lys Arg Glu Glu Leu	290	295	300
Gln Gln Ile Leu	Ser Arg Val Lys Gln Phe Glu Ala Glu Pro Asn	305	310	315
Lys Ser Gly Leu	Lys Thr Phe Gln Thr Leu Leu Asn Thr Ile Pro	320	325	330
Gly Trp Leu Ile	Ser Glu Asp Lys Arg Glu Tyr Ala Val His Ile	335	340	345
Ala Met Glu Asn	Asn Leu Glu Lys Val Lys Glu Glu Ile Thr His	350	355	360
Ile Lys Thr Gln	Ala Glu Asp Met Leu Val Ser Tyr Glu Asn Ile	365	370	375
Ile Gln Thr Ala	Met Met Ser Ser Lys Thr Gly Lys Pro Gly Asn	380	385	390
Lys Pro Thr Ser	Leu Asp Glu Thr Ser Ser Lys Val Ser Asn Val	395	400	405
His Val Ser Asn	Asn Lys Asn Ser Glu Gln Lys Glu Asn Lys Ile	410	415	420
Ala Lys Glu Lys	Thr Val Gln His Gln Val Ala Ala His His Glu	425	430	435
Ala Thr Val Arg	Ser His Val Lys Thr His Gln Glu Ile Lys Leu	440	445	450
Asp Asp Ser Asn	Ile Pro Pro Pro Ser Leu Lys Thr Arg Pro Pro	455	460	465
Ser Pro Thr Phe	Ile Thr Ile Glu Ser Thr Ala Arg Arg Thr Glu	470	475	480
Asn Pro Thr Lys	Asn Glu Leu Ser Gln Ser Pro Lys Lys Asp Ser	485	490	495
Tyr Val Glu Pro	Pro Pro Arg Arg Pro Met Ser Gln Lys Ser Glu	500	505	510
Ile His Arg Ala	Asn Thr Ser Pro Ser Pro Pro Arg Ser Arg Ser	515	520	525
Glu Gln Leu Val	Arg Leu Lys Asp Thr Thr Ala Lys Leu Ser Lys	530	535	540
Gly Ala Ile Pro	Cys Pro Ala Ala Thr Pro Val Pro Ile Val Glu	545	550	555
Lys Arg Ser Glu	Ile Ile Met Ser Pro Ala Thr Leu Arg Arg Gln	560	565	570
Ile Lys Ile Glu	Thr Arg Gly Arg Asp Ser Pro Pro Thr Ile Thr	575	580	585
Ile Pro Val Asn	Ile Asn His Ala Ala Ser Gly Ser Phe Arg Glu	590	595	600
Ser Val Asp Ala	Gln Glu Glu Ile Arg Lys Val Glu Lys Arg Ala			

Thr Tyr Val His	605	610	615
Lys Asp Gly Leu Asn Ser Thr Asp His Met Val	620	625	630
Pro Asp Thr Glu Ser Tyr Asp Ala Val Glu Ile Ile Arg Lys Val	635	640	645
Ala Val Pro Pro Arg Leu Ser Glu His Thr Gln Arg Tyr Glu Ala	650	655	660
Ala Asn Arg Thr Val Gln Met Ala Glu Asn Phe Val Asn Asp Pro	665	670	675
Glu Asn Glu Ile Asn Arg Trp Phe Arg Glu Phe Glu His Gly Pro	680	685	690
Val Ser Glu Ala Lys Ser Asn Arg Arg Val Tyr Ala Lys Gly Glu	695	700	705
Thr Asn His Asn Ile Gln Gln Glu Ser Arg Thr Phe Cys Lys Glu	710	715	720
Glu Phe Gly Leu Thr Ser Leu Gly Asn Thr Ser Phe Thr Asp Phe	725	730	735
Ser Cys Lys His Pro Arg Glu Leu Arg Glu Lys Ile Pro Val Lys	740	745	750
Gln Pro Arg Ile Cys Ser Glu Thr Arg Ser Leu Ser Glu His Phe	755	760	765
Ser Gly Met Asp Ala Phe Glu Ser Gln Ile Val Glu Ser Lys Met	770	775	780
Lys Thr Ser Ser Ser His Ser Ser Glu Ala Gly Lys Ser Gly Cys	785	790	795
Asp Phe Lys His Ala Pro Pro Thr Tyr Glu Asp Val Ile Ala Gly	800	805	810
His Ile Leu Asp Ile Ser Asp Ser Pro Lys Glu Val Arg Lys Asn	815	820	825
Phe Gln Lys Thr Trp Gln Glu Ser Gly Arg Val Phe Lys Gly Leu	830	835	840
Gly Tyr Ala Thr Ala Asp Ala Ser Ala Thr Glu Met Arg Thr Thr	845	850	855
Phe Gln Glu Glu Ser Ala Phe Ile Ser Glu Ala Ala Ala Pro Arg	860	865	870
Gln Gly Asn Met Tyr Thr Leu Ser Lys Asp Ser Leu Ser Asn Gly	875	880	885
Val Pro Ser Gly Arg Gln Ala Glu Phe Ser	890	895	

&lt;210&gt; 26

&lt;211&gt; 109

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505908CD1

&lt;400&gt; 26

Met Glu Ser Asp Phe Tyr Leu Arg Tyr Tyr Val Gly His Lys Gly	1	5	10	15
Lys Phe Gly His Glu Phe Leu Glu Phe Glu Phe Arg Pro Asp Gly	20	25	30	
Lys Leu Arg Tyr Ala Asn Asn Ser Asn Tyr Lys Asn Asp Val Met	35	40	45	
Ile Arg Lys Glu Glu Leu Glu Ile Val Ile Gly Asp Glu His Ile	50	55	60	
Ser Phe Thr Thr Ser Lys Ile Gly Ser Leu Ile Asp Val Asn Gln	65	70	75	
Ser Lys Asp Pro Glu Gly Leu Arg Val Phe Tyr Tyr Leu Val Gln	80	85	90	
Asp Leu Lys Cys Leu Val Phe Ser Leu Ile Gly Leu His Phe Lys				

Ile Lys Pro Ile 95

100

105

<210> 27  
 <211> 334  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 6590147CD1

<400> 27  
 Met Lys Phe Pro Gly Pro Leu Glu Asn Gln Arg Leu Ser Phe Leu  
 1 5 10 15  
 Leu Glu Lys Ala Ile Thr Arg Glu Ala Gln Met Trp Lys Val Asn  
 20 25 30  
 Val Arg Lys Met Pro Ser Asn Gln Ala His Pro Lys Tyr Leu Ser  
 35 40 45  
 Cys Ile Ala Ile Ser Cys Phe Phe Leu Ala Ala Lys Thr Val Glu  
 50 55 60  
 Glu Asp Glu Arg Ile Pro Val Leu Lys Val Leu Ala Arg Asp Ser  
 65 70 75  
 Phe Cys Gly Cys Ser Ser Ser Glu Ile Leu Arg Met Glu Arg Ile  
 80 85 90  
 Ile Leu Asp Lys Leu Asn Trp Asp Leu His Thr Ala Thr Pro Leu  
 95 100 105  
 Asp Phe Leu His Ile Phe His Ala Ile Ala Val Ser Thr Arg Pro  
 110 115 120  
 Gln Leu Leu Phe Ser Leu Pro Lys Leu Ser Pro Ser Gln His Leu  
 125 130 135  
 Ala Val Leu Thr Lys Gln Leu Leu His Cys Met Ala Cys Asn Gln  
 140 145 150  
 Leu Leu Gln Phe Arg Gly Ser Met Leu Ala Leu Ala Met Val Ser  
 155 160 165  
 Leu Glu Met Glu Lys Leu Ile Pro Asp Trp Leu Ser Leu Thr Ile  
 170 175 180  
 Glu Leu Leu Gln Lys Ala Gln Met Asp Ser Ser Gln Leu Ile His  
 185 190 195  
 Cys Arg Glu Leu Val Ala His His Leu Ser Thr Leu Gln Ser Ser  
 200 205 210  
 Leu Pro Leu Asn Ser Val Tyr Val Tyr Arg Pro Leu Lys His Thr  
 215 220 225  
 Leu Val Thr Cys Asp Lys Gly Val Phe Arg Leu His Pro Ser Ser  
 230 235 240  
 Val Pro Gly Pro Asp Phe Ser Lys Asp Asn Ser Lys Pro Glu Val  
 245 250 255  
 Pro Val Arg Gly Thr Ala Ala Phe Tyr His His Leu Pro Ala Ala  
 260 265 270  
 Ser Gly Cys Lys Gln Thr Ser Thr Lys Arg Lys Val Glu Glu Met  
 275 280 285  
 Glu Val Asp Asp Phe Tyr Asp Gly Ile Lys Arg Leu Tyr Asn Glu  
 290 295 300  
 Asp Asn Val Ser Glu Asn Val Gly Ser Val Cys Gly Thr Asp Leu  
 305 310 315  
 Ser Arg Gln Glu Gly His Ala Ser Pro Cys Pro Pro Leu Gln Pro  
 320 325 330  
 Val Ser Val Met

<210> 28  
 <211> 569

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6828539CD1

&lt;400&gt; 28

```

Met Ser Thr Ala Thr Gly Pro Glu Ala Ala Pro Lys Pro Ser Ala
 1          5          10          15
Lys Ser Ile Tyr Glu Gln Arg Lys Arg Tyr Ser Thr Val Val Met
          20          25          30
Ala Asp Val Ser Gln Tyr Pro Val Asn His Leu Val Thr Phe Cys
          35          40          45
Leu Gly Glu Asp Asp Gly Val His Thr Val Glu Asp Ala Ser Arg
          50          55          60
Lys Leu Ala Val Met Asp Ser Gln Gly Arg Val Trp Ala Gln Glu
          65          70          75
Met Leu Leu Arg Val Ser Pro Asp His Val Thr Leu Leu Asp Pro
          80          85          90
Ala Ser Lys Glu Glu Leu Glu Ser Tyr Pro Leu Gly Ala Ile Val
          95          100          105
Arg Cys Asp Ala Val Met Pro Pro Gly Arg Ser Arg Ser Leu Leu
          110          115          120
Leu Leu Val Cys Gln Glu Pro Glu Arg Ala Gln Pro Asp Val His
          125          130          135
Phe Phe Gln Gly Leu Arg Leu Gly Ala Glu Leu Ile Arg Glu Asp
          140          145          150
Ile Gln Gly Ala Leu His Asn Tyr Arg Ser Gly Arg Gly Glu Arg
          155          160          165
Arg Ala Ala Ala Leu Arg Ala Thr Gln Glu Glu Leu Gln Arg Asp
          170          175          180
Arg Ser Pro Ala Ala Glu Thr Pro Pro Leu Gln Arg Arg Pro Ser
          185          190          195
Val Arg Ala Val Ile Ser Thr Val Glu Arg Gly Ala Gly Arg Gly
          200          205          210
Arg Pro Gln Ala Lys Pro Ile Pro Glu Ala Glu Glu Ala Gln Arg
          215          220          225
Pro Glu Pro Val Gly Thr Ser Ser Asn Ala Asp Ser Ala Ser Pro
          230          235          240
Asp Leu Gly Pro Arg Gly Pro Asp Leu Ala Val Leu Gln Ala Glu
          245          250          255
Arg Glu Val Asp Ile Leu Asn His Val Phe Asp Asp Val Glu Ser
          260          265          270
Phe Val Ser Arg Leu Gln Lys Ser Ala Glu Ala Ala Arg Val Leu
          275          280          285
Glu His Arg Glu Arg Gly Arg Arg Ser Arg Arg Arg Ala Ala Gly
          290          295          300
Glu Gly Leu Leu Thr Leu Arg Ala Lys Pro Pro Ser Glu Ala Glu
          305          310          315
Tyr Thr Asp Val Leu Gln Lys Ile Lys Tyr Ala Phe Ser Leu Leu
          320          325          330
Ala Arg Leu Arg Gly Asn Ile Ala Asp Pro Ser Ser Pro Glu Leu
          335          340          345
Leu His Phe Leu Phe Gly Pro Leu Gln Met Ile Val Asn Thr Ser
          350          355          360
Gly Gly Pro Glu Phe Ala Ser Ser Val Arg Arg Pro His Leu Thr
          365          370          375
Ser Asp Ala Val Ala Leu Leu Arg Asp Asn Val Thr Pro Arg Glu
          380          385          390
Asn Glu Leu Trp Thr Ser Leu Gly Asp Ser Trp Thr Arg Pro Gly
          395          400          405
Leu Glu Leu Ser Pro Glu Glu Gly Pro Pro Tyr Arg Pro Glu Phe

```

Phe Ser Gly Trp	410	415	420
Glu Pro Pro Val Thr	425	430	435
Trp Glu Asp Pro	440	445	450
Arg Gln Val Thr	455	460	465
Arg Gly Arg Gly	470	475	480
Tyr Phe Leu Gln	485	490	495
Arg Asp Leu Glu	500	505	510
Ala Gly Lys Trp	515	520	525
Ser Ser Glu Leu	530	535	540
Glu Asp Lys Glu	545	550	555
Gln Lys Lys Lys	560	565	

&lt;210&gt; 29

&lt;211&gt; 429

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7170321CD1

&lt;400&gt; 29

Met Val Asp Tyr Ser Val Trp Asp His Ile Glu Val Ser Asp Asp	1	5	10	15
Glu Asp Glu Thr His Pro Asn Ile Asp Thr Ala Ser Leu Phe Arg	20	25	30	35
Trp Arg His Gln Ala Arg Val Glu Arg Met Glu Gln Phe Gln Lys	40	45	50	55
Glu Lys Glu Glu Leu Asp Arg Gly Cys Arg Glu Cys Lys Arg Lys	60	65	70	75
Val Ala Glu Cys Gln Arg Lys Leu Lys Glu Leu Glu Val Ala Glu	80	85	90	95
Gly Gly Lys Ala Glu Leu Glu Arg Leu Gln Ala Glu Ala Gln Gln	100	105	110	115
Leu Arg Lys Glu Glu Arg Ser Trp Glu Gln Lys Leu Glu Glu Met	120	125	130	135
Arg Lys Lys Glu Lys Ser Met Pro Trp Asn Val Asp Thr Leu Ser	140	145	150	155
Lys Asp Gly Phe Ser Lys Ala His Ala His Ala Pro Ala Thr Leu	160	165	170	175
Met Leu Met Ala Ser Ser Pro Pro Pro Pro Trp Met Gly Trp Ala	180	185	190	195
Leu Arg Pro Arg Pro Ala Arg Pro Arg Ser Arg Arg Ser Ala Ala	200	205	210	215
Leu Pro Gln Pro Thr Pro Pro Leu Thr Leu Pro Gln Ser Met Val	220	225	230	235
Asn Thr Lys Pro Glu Lys Thr Glu Glu Asp Ser Glu Glu Val Arg	240	245	250	255
Glu Gln Lys His Lys Thr Phe Val Glu Lys Tyr Glu Lys Gln Ile	260	265	270	275
Lys His Phe Gly Met Leu Arg Arg Trp Asp Asp Ser Gln Lys Tyr	280	285	290	295
Leu Ser Asp Asn Val His Leu Val Cys Glu Glu Thr Ala Asn Tyr	300	305	310	315



Leu Val Ile Trp	230	Cys Ile Asp Leu Glu	235	Val Glu Glu Lys Cys	240
245	250	255	260	265	270
Leu Met Glu Gln	245	Val Ala His Gln Thr	250	Ile Val Met Gln Phe	255
260	265	270	275	280	285
Leu Glu Leu Ala	275	Lys Ser Leu Lys Val	280	Asp Pro Arg Ala Cys	285
290	295	300	305	310	315
Arg Gln Phe Phe	290	Thr Lys Ile Lys Thr	295	Ala Asp Arg Gln Tyr	300
310	315	320	325	330	335
Glu Gly Phe Asn	305	Asp Glu Leu Glu Ala	310	Phe Lys Glu Arg Val	315
320	325	330	335	340	345
Gly Arg Ala Lys	320	Leu Arg Ile Glu Lys	325	Ala Met Lys Glu Tyr	330
335	340	345	350	355	360
Glu Glu Glu Arg	335	Lys Lys Arg Leu Gly	340	Pro Gly Gly Leu Asp	345
350	355	360	365	370	375
Val Glu Val Tyr	350	Glu Ser Leu Pro Glu	355	Glu Leu Gln Lys Cys	360
365	370	375	380	385	390
Asp Val Lys Asp	365	Val Gln Met Leu Gln	370	Asp Ala Ile Ser Lys	375
375	380	385	390	395	400
Asp Pro Thr Asp	380	Ala Lys Tyr His Met	385	Gln Arg Cys Ile Asp	390
395	400	405	410	415	420
Gly Leu Trp Val	395	Pro Asn Ser Lys Ala	400	Ser Glu Ala Lys Glu	405
405	410	415	420	425	
Glu Glu Ala Gly	405	Pro Gly Asp Pro Leu	410	Leu Glu Ala Val Pro	415
415	420	425			
Thr Gly Asp Glu	415	Lys Asp Val Ser Val	420		

&lt;210&gt; 30

&lt;211&gt; 776

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505918CD1

&lt;400&gt; 30

Met Asp Thr Glu Gly	5	Phe Gly Glu Leu Leu	10	Gln Gln Ala Glu Gln	15
1	10	15	20	25	30
Leu Ala Ala Glu Thr	20	Glu Gly Ile Ser	25	Glu Leu Pro His Val	30
30	35	40	45	50	55
Arg Asn Leu Gln Glu	35	Ile Gln Gln Ala Gly	40	Glu Arg Leu Arg Ser	45
45	50	55	60	65	70
Arg Thr Leu Thr Arg	50	Thr Ser Gln Glu Thr	55	Ala Asp Val Lys Ala	60
60	65	70	75	80	85
Ser Val Leu Leu Gly	65	Ser Arg Gly Leu Asp	70	Ile Ser His Ile Ser	75
75	80	85	90	95	100
Gln Arg Leu Glu Ser	80	Leu Ser Ala Ala Thr	85	Thr Phe Glu Pro Leu	90
90	95	100	105	110	115
Glu Pro Val Lys Asp	95	Thr Asp Ile Gln Gly	100	Phe Leu Lys Asn Glu	105
105	110	115	120	125	130
Lys Asp Asn Ala Leu	110	Leu Ser Ala Ile Glu	115	Glu Ser Arg Lys Arg	120
120	125	130	135	140	145
Thr Phe Gly Met Ala	125	Glu Glu Tyr His Arg	130	Glu Ser Met Leu Val	135
135	140	145	150	155	160
Glu Trp Glu Gln Val	140	Lys Gln Arg Ile Leu	145	His Thr Leu Leu Ala	150
150	155	160	165	170	175
Ser Gly Glu Asp Ala	155	Leu Asp Phe Thr Gln	160	Glu Ser Glu Pro Ser	165
165	170	175	180	185	190
Tyr Ile Ser Asp Val	170	Gly Pro Pro Gly Arg	175	Ser Ser Leu Asp Asn	180
180	185	190	195	200	205
Ile Glu Met Ala Tyr	185	Ala Arg Gln Ile Tyr	190	Ile Tyr Asn Glu Lys	195

	185		190		195
Ile Val Asn Gly	His Leu Gln Pro Asn	Leu Val Asp Leu Cys	Ala		
	200		205		210
Ser Val Ala Glu	Leu Asp Asp Lys Ser	Ile Ser Asp Met Trp	Thr		
	215		220		225
Met Val Lys Gln	Met Thr Asp Val Leu	Leu Thr Pro Ala Thr	Asp		
	230		235		240
Ala Leu Lys Asn	Arg Ser Ser Val Glu	Val Arg Met Glu Phe	Val		
	245		250		255
Arg Gln Ala Leu	Ala Tyr Leu Glu Gln	Ser Tyr Lys Asn Tyr	Thr		
	260		265		270
Leu Val Thr Val	Phe Gly Asn Leu His	Gln Ala Gln Leu Gly	Gly		
	275		280		285
Val Pro Gly Thr	Tyr Gln Leu Val Arg	Ser Phe Leu Asn Ile	Lys		
	290		295		300
Leu Pro Ala Pro	Leu Pro Gly Leu Gln	Asp Gly Glu Val Glu	Gly		
	305		310		315
His Pro Val Trp	Ala Leu Ile Tyr Tyr	Cys Met Arg Cys Gly	Asp		
	320		325		330
Leu Leu Ala Ala	Ser Gln Val Val Asn	Arg Ala Gln His Gln	Leu		
	335		340		345
Gly Glu Phe Lys	Thr Trp Phe Gln Glu	Tyr Met Asn Ser Lys	Asp		
	350		355		360
Arg Arg Leu Ser	Pro Ala Thr Glu Asn	Lys Leu Arg Leu His	Tyr		
	365		370		375
Arg Arg Ala Leu	Arg Asn Asn Thr Asp	Pro Tyr Lys Arg Ala	Val		
	380		385		390
Tyr Cys Ile Ile	Gly Arg Cys Asp Val	Thr Asp Asn Gln Ser	Glu		
	395		400		405
Val Ala Asp Lys	Thr Glu Asp Tyr Leu	Trp Leu Lys Leu Asn	Gln		
	410		415		420
Val Cys Phe Asp	Asp Asp Gly Thr Ser	Ser Pro Gln Asp Arg	Leu		
	425		430		435
Thr Leu Ser Gln	Phe Gln Lys Gln Leu	Leu Glu Asp Tyr Gly	Glu		
	440		445		450
Ser His Phe Thr	Val Asn Gln Gln Pro	Phe Leu Tyr Phe Gln	Val		
	455		460		465
Leu Phe Leu Thr	Ala Gln Phe Glu Ala	Ala Val Ala Phe Leu	Phe		
	470		475		480
Arg Met Glu Arg	Leu Arg Cys His Ala	Val His Val Ala Leu	Val		
	485		490		495
Leu Phe Glu Leu	Lys Leu Leu Leu Lys	Ser Ser Gly Gln Ser	Ala		
	500		505		510
Gln Leu Leu Ser	His Glu Pro Gly Asp	Pro Pro Cys Leu Arg	Arg		
	515		520		525
Leu Asn Phe Val	Arg Leu Leu Met Leu	Tyr Thr Arg Lys Phe	Glu		
	530		535		540
Ser Thr Asp Pro	Arg Glu Ala Leu Gln	Tyr Phe Tyr Phe Leu	Arg		
	545		550		555
Asp Glu Lys Asp	Ser Gln Gly Glu Asn	Met Phe Leu Arg Cys	Val		
	560		565		570
Ser Glu Leu Val	Ile Glu Ser Arg Glu	Phe Asp Met Ile Leu	Gly		
	575		580		585
Lys Leu Glu Asn	Asp Gly Ser Arg Lys	Pro Gly Val Ile Asp	Lys		
	590		595		600
Phe Thr Ser Asp	Thr Lys Pro Ile Ile	Asn Lys Val Ala Ser	Val		
	605		610		615
Ala Glu Asn Lys	Gly Leu Phe Glu Glu	Ala Ala Lys Leu Tyr	Asp		
	620		625		630
Leu Ala Lys Asn	Ala Asp Lys Val Leu	Glu Leu Met Asn Lys	Leu		
	635		640		645
Leu Ser Pro Val	Val Pro Gln Ile Ser	Ala Pro Gln Ser Asn	Lys		
	650		655		660

Glu Arg Leu Lys	Asn Met Ala Leu Ser Ile Ala Glu Arg Tyr Arg	
	665	670
Ala Gln Gly Ile	Ser Ala Asn Lys Phe Val Asp Ser Thr Phe Tyr	675
	680	685
Leu Leu Leu Asp	Leu Ile Thr Phe Phe Asp Glu Tyr His Ser Gly	690
	695	700
His Ile Asp Arg	Ala Phe Asp Ile Ile Glu Arg Leu Lys Leu Val	705
	710	715
Pro Leu Asn Gln	Glu Ser Val Glu Glu Arg Val Ala Ala Phe Arg	720
	725	730
Asn Phe Ser Asp	Glu Gln Leu Arg Ser Gln Ala Arg Thr Leu Ile	735
	740	745
Thr Phe Ala Gly	Met Ile Pro Tyr Arg Thr Ser Gly Asp Thr Asn	750
	755	760
Ala Arg Leu Val	Gln Met Glu Val Leu Met Asn	765
	770	775

&lt;210&gt; 31

&lt;211&gt; 975

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505935CD1

&lt;400&gt; 31

Met Met Ala Gln Ser Asn Met Phe Thr Val Ala Asp Val Leu Ser	
1	5
Gln Asp Glu Leu Arg Lys Lys Leu Tyr Gln Thr Phe Lys Asp Arg	10
	15
Gly Ile Leu Asp Thr Leu Lys Thr Gln Leu Arg Asn Gln Leu Ile	20
	25
His Glu Leu Met His Pro Val Leu Ser Gly Glu Leu Gln Pro Arg	30
	35
Ser Ile Ser Val Glu Gly Ser Ser Leu Leu Ile Gly Ala Ser Asn	40
	45
Ser Leu Val Ala Asp His Leu Gln Arg Cys Gly Tyr Glu Tyr Ser	50
	55
Leu Ser Val Phe Phe Pro Glu Ser Gly Leu Ala Lys Glu Lys Val	60
	65
Phe Thr Met Gln Asp Leu Leu Gln Leu Ile Lys Ile Asn Pro Thr	70
	75
Ser Ser Leu Tyr Lys Ser Leu Val Ser Gly Ser Asp Lys Glu Asn	80
	85
Gln Lys Gly Phe Leu Met His Phe Leu Lys Glu Leu Ala Glu Tyr	90
	95
His Gln Ala Lys Glu Ser Cys Asn Met Glu Thr Gln Thr Ser Ser	100
	105
Thr Phe Asn Arg Asp Ser Leu Ala Glu Lys Leu Gln Leu Ile Asp	110
	115
Asp Gln Phe Ala Asp Ala Tyr Pro Gln Arg Ile Lys Phe Glu Ser	120
	125
Leu Glu Ile Lys Leu Asn Glu Tyr Lys Arg Glu Ile Glu Glu Gln	130
	135
Leu Arg Ala Glu Met Cys Gln Lys Leu Lys Phe Phe Lys Asp Thr	140
	145
Glu Ile Ala Lys Ile Lys Met Glu Ala Lys Lys Lys Tyr Glu Lys	150
	155
Glu Leu Thr Met Phe Gln Asn Asp Phe Glu Lys Ala Cys Gln Ala	160
	165
Lys Ser Glu Ala Leu Val Leu Arg Glu Lys Ser Thr Leu Glu Arg	170
	175
	180
	185
	190
	195
	200
	205
	210
	215
	220
	225
	230
	235
	240
	245
	250
	255
	260
	265
	270

Ile His Lys His Gln Glu Ile Glu Thr Lys Glu Ile Tyr Ala Gln  
 275 280 285  
 Arg Gln Leu Leu Leu Lys Asp Met Asp Leu Leu Arg Gly Arg Glu  
 290 295 300  
 Ala Glu Leu Lys Gln Arg Val Glu Ala Phe Glu Leu Asn Gln Lys  
 305 310 315  
 Leu Gln Glu Glu Lys His Lys Ser Ile Thr Glu Ala Leu Arg Arg  
 320 325 330  
 Gln Glu Gln Asn Ile Lys Ser Phe Glu Glu Thr Tyr Asp Arg Lys  
 335 340 345  
 Leu Lys Asn Glu Leu Leu Lys Tyr Gln Leu Glu Leu Lys Asp Asp  
 350 355 360  
 Tyr Ile Ile Arg Thr Asn Arg Leu Ile Glu Asp Glu Arg Lys Asn  
 365 370 375  
 Lys Glu Lys Ala Val His Leu Gln Glu Glu Leu Ile Ala Ile Asn  
 380 385 390  
 Ser Lys Lys Glu Glu Leu Asn Gln Ser Val Asn Arg Val Lys Glu  
 395 400 405  
 Leu Glu Leu Glu Leu Glu Ser Val Lys Ala Gln Ser Leu Ala Ile  
 410 415 420  
 Thr Lys Gln Asn His Met Leu Asn Glu Lys Val Lys Glu Met Ser  
 425 430 435  
 Asp Tyr Ser Leu Leu Lys Glu Glu Lys Leu Glu Leu Leu Ala Gln  
 440 445 450  
 Asn Lys Leu Leu Lys Gln Gln Leu Glu Glu Ser Arg Asn Glu Asn  
 455 460 465  
 Leu Arg Leu Leu Asn Arg Leu Ala Gln Pro Ala Pro Glu Leu Ala  
 470 475 480  
 Val Phe Gln Lys Glu Leu Arg Lys Ala Glu Lys Ala Ile Val Val  
 485 490 495  
 Glu His Glu Glu Phe Glu Ser Cys Arg Gln Ala Leu His Lys Gln  
 500 505 510  
 Leu Gln Asp Glu Ile Glu His Ser Ala Gln Leu Lys Ala Gln Ile  
 515 520 525  
 Leu Gly Tyr Lys Ala Ser Val Lys Ser Leu Thr Thr Gln Val Ala  
 530 535 540  
 Asp Leu Lys Leu Gln Leu Lys Gln Thr Gln Thr Ala Leu Glu Asn  
 545 550 555  
 Glu Val Tyr Cys Asn Pro Lys Gln Ser Val Ile Asp Arg Ser Val  
 560 565 570  
 Asn Gly Leu Ile Asn Gly Asn Val Val Pro Cys Asn Gly Glu Ile  
 575 580 585  
 Ser Gly Asp Phe Leu Asn Asn Pro Phe Lys Gln Glu Asn Val Leu  
 590 595 600  
 Ala Arg Met Val Ala Ser Arg Ile Thr Asn Tyr Pro Thr Ala Trp  
 605 610 615  
 Val Glu Gly Ser Ser Pro Asp Ser Asp Leu Glu Phe Val Ala Asn  
 620 625 630  
 Thr Lys Ala Arg Val Lys Glu Leu Gln Gln Glu Ala Glu Arg Leu  
 635 640 645  
 Glu Lys Ala Phe Arg Ser Tyr His Arg Arg Val Ile Lys Asn Ser  
 650 655 660  
 Ala Lys Ser Pro Leu Ala Ala Lys Ser Pro Pro Ser Leu His Leu  
 665 670 675  
 Leu Glu Ala Phe Lys Asn Ile Thr Ser Ser Ser Pro Glu Arg His  
 680 685 690  
 Ile Phe Gly Glu Asp Arg Val Val Ser Glu Gln Pro Gln Val Gly  
 695 700 705  
 Thr Leu Glu Glu Arg Asn Asp Val Val Glu Ala Leu Thr Gly Ser  
 710 715 720  
 Ala Ala Ser Arg Leu Arg Gly Gly Thr Ser Ser Arg Arg Leu Ser  
 725 730 735  
 Ser Thr Pro Leu Pro Lys Ala Lys Arg Ser Leu Glu Ser Glu Met

	740		745		750
Tyr Leu Glu Gly	Leu Gly Arg Ser His	Ile Ala Ser Pro Ser	Pro		
	755		760		765
Cys Pro Asp Arg	Met Pro Leu Pro Ser	Pro Thr Glu Ser Arg	His		
	770		775		780
Ser Leu Ser Ile	Pro Pro Val Ser Ser	Pro Pro Glu Gln Lys	Val		
	785		790		795
Gly Leu Tyr Arg	Arg Gln Thr Glu Leu	Gln Asp Lys Ser Glu	Phe		
	800		805		810
Ser Asp Val Asp	Lys Leu Ala Phe Lys	Asp Asn Glu Glu Phe	Glu		
	815		820		825
Ser Ser Phe Glu	Cys Val Asp Gln Lys	Gln Ile Glu Glu Gln	Lys		
	830		835		840
Glu Glu Glu Lys	Ile Arg Glu Gln Gln	Val Lys Glu Arg Arg	Gln		
	845		850		855
Arg Glu Glu Arg	Arg Gln Ser Asn Leu	Gln Glu Val Leu Glu	Arg		
	860		865		870
Glu Arg Arg Glu	Leu Glu Lys Leu Tyr	Gln Glu Arg Lys Met	Ile		
	875		880		885
Glu Glu Ser Leu	Lys Ile Lys Ile Lys	Lys Glu Leu Glu Met	Glu		
	890		895		900
Asn Glu Leu Glu	Met Ser Asn Gln Glu	Ile Lys Asp Lys Ser	Ala		
	905		910		915
His Ser Glu Asn	Pro Leu Glu Lys Tyr	Met Lys Ile Ile Gln	Gln		
	920		925		930
Glu Gln Asp Gln	Glu Ser Ala Asp Lys	Ser Ser Lys Lys Met	Val		
	935		940		945
Gln Glu Gly Ser	Leu Val Asp Thr Leu	Gln Ser Ser Asp Lys	Val		
	950		955		960
Glu Ser Leu Thr	Gly Phe Ser His Glu	Glu Leu Asp Asp Ser	Trp		
	965		970		975

&lt;210&gt; 32

&lt;211&gt; 814

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4225965CD1

&lt;400&gt; 32

Met Asp Pro Gln Pro	Leu Arg Gly Ala Ser	Glu Glu Pro Ser Gly
1	5	10
Thr Gln Ser Glu Gly	Gly Gly Ser Ser Ser	Ser Gly Ala Gly Ser
20	25	30
Pro Gly Pro Pro Gly	Ile Leu Arg Pro Leu	Gln Pro Pro Gln Arg
35	40	45
Ala Asp Thr Pro Arg	Arg Asn Ser Ser Ser	Ser Ser Ser Pro Ser
50	55	60
Glu Trp Pro Arg Gln	Lys Leu Ser Arg Lys	Ala Ile Ser Ser Ala
65	70	75
Asn Leu Leu Val Arg	Ser Gly Ser Thr Glu	Ser Arg Gly Gly Lys
80	85	90
Asp Pro Leu Ser Ser	Pro Gly Gly Pro Gly	Ser Arg Arg Ser Asn
95	100	105
Tyr Asn Leu Glu Gly	Ile Ser Val Lys Met	Phe Leu Arg Gly Arg
110	115	120
Pro Ile Thr Met Tyr	Ile Pro Ser Gly Ile	Arg Ser Leu Glu Glu
125	130	135
Leu Pro Ser Gly Pro	Pro Pro Glu Thr Leu	Ser Leu Asp Trp Val
140	145	150

Tyr Gly Tyr Arg	Gly Arg Asp Ser Arg Ser Asn Leu Phe Val Leu	155	160	165
Arg Ser Gly Glu	Val Val Tyr Phe Ile Ala Cys Val Val Val Leu	170	175	180
Tyr Arg Pro Gly	Gly Gly Pro Gly Gly Pro Gly Gly Gly Gly Gln	185	190	195
Arg His Tyr Arg	Gly His Thr Asp Cys Val Arg Cys Leu Ala Val	200	205	210
His Pro Asp Gly	Val Arg Val Ala Ser Gly Gln Thr Ala Gly Val	215	220	225
Asp Lys Asp Gly	Lys Pro Leu Gln Pro Val Val His Ile Trp Asp	230	235	240
Ser Glu Thr Leu	Leu Lys Leu Gln Glu Ile Gly Leu Gly Ala Phe	245	250	255
Glu Arg Gly Val	Gly Ala Leu Ala Phe Ser Ala Ala Asp Gln Gly	260	265	270
Ala Phe Leu Cys	Val Val Asp Asp Ser Asn Glu His Met Leu Ser	275	280	285
Val Trp Asp Cys	Ser Arg Gly Met Lys Leu Ala Glu Ile Lys Ser	290	295	300
Thr Asn Asp Ser	Val Leu Ala Val Gly Phe Asn Pro Arg Asp Ser	305	310	315
Ser Cys Ile Val	Thr Ser Gly Lys Ser His Val His Phe Trp Asn	320	325	330
Trp Ser Gly Gly	Val Gly Val Pro Gly Asn Gly Thr Leu Thr Arg	335	340	345
Lys Gln Gly Val	Phe Gly Lys Tyr Lys Lys Pro Lys Phe Ile Pro	350	355	360
Cys Phe Val Phe	Leu Pro Asp Gly Asp Ile Leu Thr Gly Asp Ser	365	370	375
Glu Gly Asn Ile	Leu Thr Trp Gly Arg Ser Pro Ser Asp Ser Lys	380	385	390
Thr Pro Gly Arg	Gly Gly Ala Lys Glu Thr Tyr Gly Ile Val Ala	395	400	405
Gln Ala His Ala	His Glu Gly Ser Ile Phe Ala Leu Cys Leu Arg	410	415	420
Arg Asp Gly Thr	Val Leu Ser Gly Gly Gly Arg Asp Arg Arg Leu	425	430	435
Val Gln Trp Gly	Pro Gly Leu Val Ala Leu Gln Glu Ala Glu Ile	440	445	450
Pro Glu His Phe	Gly Ala Val Arg Ala Ile Ala Glu Gly Leu Gly	455	460	465
Ser Glu Leu Leu	Val Gly Thr Thr Lys Asn Ala Leu Leu Arg Gly	470	475	480
Asp Leu Ala Gln	Gly Phe Ser Pro Val Ile Gln Gly His Thr Asp	485	490	495
Glu Leu Trp Gly	Leu Cys Thr His Pro Ser Gln Asn Arg Phe Leu	500	505	510
Thr Cys Gly His	Asp Arg Gln Leu Cys Leu Trp Asp Gly Glu Ser	515	520	525
His Ala Leu Ala	Trp Ser Ile Asp Leu Lys Glu Thr Gly Leu Cys	530	535	540
Ala Asp Phe His	Pro Ser Gly Ala Val Val Ala Val Gly Leu Asn	545	550	555
Thr Gly Arg Trp	Leu Val Leu Asp Thr Glu Thr Arg Glu Ile Val	560	565	570
Ser Asp Val Ile	Asp Gly Asn Glu Gln Leu Ser Val Val Arg Tyr	575	580	585
Ser Pro Asp Gly	Leu Tyr Leu Ala Ile Gly Ser His Asp Asn Val	590	595	600
Ile Tyr Ile Tyr	Ser Val Ser Ser Asp Gly Ala Lys Ser Ser Arg	605	610	615
Phe Gly Arg Cys	Met Gly His Ser Ser Phe Ile Thr His Leu Asp			

620	625	630
Trp Ser Lys Asp Gly Asn Phe Ile Met	Ser Asn Ser Gly Asp Tyr	
635	640	645
Glu Ile Leu Tyr Trp Asp Val Ala Gly	Gly Cys Lys Gln Leu Lys	
650	655	660
Asn Arg Tyr Glu Ser Arg Asp Arg Glu	Trp Ala Thr Tyr Thr Cys	
665	670	675
Val Leu Gly Phe His Val Tyr Val Pro	Val Arg Ser Cys Gln Gly	
680	685	690
Ala Glu Pro His Val Arg Gly Pro Arg	Gln Pro Arg Asp Gln Arg	
695	700	705
Pro Ile His Ala Arg Arg Leu Ala Pro	Arg Leu Ala Gly Arg Gln	
710	715	720
Gly Arg Gln His Leu Pro Val Ala Ser	Ala Gly Arg Trp Gly Arg	
725	730	735
Gly Ala Gly Ala Arg His Ala Leu Ser	Asn Pro Leu Pro Val Pro	
740	745	750
Arg Leu Leu Pro Arg Arg Leu Ile Ala	Ala Trp Arg Asp Arg Leu	
755	760	765
Ala Arg Arg Arg Gly Pro Ala Pro Pro	Cys Pro Ser Leu Ala Gln	
770	775	780
Ser Pro Thr Thr Arg Gly Arg Leu Phe	Pro Gly Leu Thr Ser Arg	
785	790	795
His Ser Arg Ser Arg Ile Phe Leu Glu	Gly Ala Asn Gly Ala Pro	
800	805	810
Ala His Thr Val		

<210> 33  
 <211> 3570  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7495594CD1

<400> 33

Met Trp Pro Arg Leu Ala Phe Cys Cys Trp Gly Leu Ala Leu Val	
1 5 10 15	
Ser Gly Trp Ala Thr Phe Gln Gln Met Ser Pro Ser Arg Asn Phe	
20 25 30	
Ser Phe Arg Leu Phe Pro Glu Thr Ala Pro Gly Ala Pro Gly Ser	
35 40 45	
Ile Pro Ala Pro Pro Ala Pro Gly Asp Glu Ala Ala Gly Ser Arg	
50 55 60	
Val Glu Arg Leu Gly Gln Ala Phe Arg Arg Val Arg Leu Leu	
65 70 75	
Arg Glu Leu Ser Glu Arg Leu Glu Leu Val Phe Leu Val Asp Asp	
80 85 90	
Ser Ser Ser Val Gly Glu Val Asn Phe Arg Ser Glu Leu Met Phe	
95 100 105	
Val Arg Lys Leu Leu Ser Asp Phe Pro Val Val Pro Thr Ala Thr	
110 115 120	
Arg Val Ala Ile Val Thr Phe Ser Ser Lys Asn Tyr Val Val Pro	
125 130 135	
Arg Val Asp Tyr Ile Ser Thr Arg Arg Ala Arg Gln His Lys Cys	
140 145 150	
Ala Leu Leu Leu Gln Glu Ile Pro Ala Ile Ser Tyr Arg Gly Gly	
155 160 165	
Gly Thr Tyr Thr Lys Gly Ala Phe Gln Gln Ala Ala Gln Ile Leu	
170 175 180	
Leu His Ala Arg Glu Asn Ser Thr Lys Val Val Phe Leu Ile Thr	





	185	190	195
Asp Gly Tyr Ser	Asn Gly Gly Asp Pro	Arg Pro Ile Ala Ala	Ser
	200	205	210
Leu Arg Asp Ser	Gly Val Glu Ile Phe	Thr Phe Gly Ile Trp	Gln
	215	220	225
Gly Asn Ile Arg	Glu Leu Asn Asp Met	Ala Ser Thr Pro Lys	Glu
	230	235	240
Glu His Cys Tyr	Leu Leu His Ser Phe	Glu Glu Phe Glu Ala	Leu
	245	250	255
Ala Arg Arg Ala	Leu His Glu Asp Leu	Pro Ser Gly Ser Phe	Ile
	260	265	270
Gln Asp Asp Met	Val His Cys Ser Tyr	Leu Cys Asp Glu Gly	Lys
	275	280	285
Asp Cys Cys Asp	Arg Met Gly Ser Cys	Lys Cys Gly Thr His	Thr
	290	295	300
Gly His Phe Glu	Cys Ile Cys Glu Lys	Gly Tyr Tyr Gly Lys	Gly
	305	310	315
Leu Gln Tyr Glu	Cys Thr Ala Cys Pro	Ser Gly Thr Tyr Lys	Pro
	320	325	330
Glu Gly Ser Pro	Gly Gly Ile Ser Ser	Cys Ile Pro Cys Pro	Asp
	335	340	345
Glu Asn His Thr	Ser Pro Pro Gly Ser	Thr Ser Pro Glu Asp	Cys
	350	355	360
Val Cys Arg Glu	Gly Tyr Arg Ala Ser	Gly Gln Thr Cys Glu	Leu
	365	370	375
Val His Cys Pro	Ala Leu Lys Pro Pro	Glu Asn Gly Tyr Phe	Ile
	380	385	390
Gln Asn Thr Cys	Asn Asn His Phe Asn	Ala Ala Cys Gly Val	Arg
	395	400	405
Cys His Pro Gly	Phe Asp Leu Val Gly	Ser Ser Ile Ile Leu	Cys
	410	415	420
Leu Pro Asn Gly	Leu Trp Ser Gly Ser	Glu Ser Tyr Cys Arg	Val
	425	430	435
Arg Thr Cys Pro	His Leu Arg Gln Pro	Lys His Gly His Ile	Ser
	440	445	450
Cys Ser Thr Arg	Glu Met Leu Tyr Lys	Thr Thr Cys Leu Val	Ala
	455	460	465
Cys Asp Glu Gly	Tyr Arg Leu Glu Gly	Ser Asp Lys Leu Thr	Cys
	470	475	480
Gln Gly Asn Ser	Gln Trp Asp Gly Pro	Glu Pro Arg Cys Val	Glu
	485	490	495
Arg His Cys Ser	Thr Phe Gln Met Pro	Lys Asp Val Ile Ile	Ser
	500	505	510
Pro His Asn Cys	Gly Lys Gln Pro Ala	Lys Phe Gly Thr Ile	Cys
	515	520	525
Tyr Val Ser Cys	Arg Gln Gly Phe Ile	Leu Ser Gly Val Lys	Glu
	530	535	540
Met Leu Arg Cys	Thr Thr Ser Gly Lys	Trp Asn Val Gly Val	Gln
	545	550	555
Ala Ala Val Cys	Lys Asp Val Glu Ala	Pro Gln Ile Asn Cys	Pro
	560	565	570
Lys Asp Ile Glu	Ala Lys Thr Leu Glu	Gln Gln Asp Ser Ala	Asn
	575	580	585
Val Thr Trp Gln	Ile Pro Thr Ala Lys	Asp Asn Ser Gly Glu	Lys
	590	595	600
Val Ser Val His	Val His Pro Ala Phe	Thr Pro Pro Tyr Leu	Phe
	605	610	615
Pro Ile Gly Asp	Val Ala Ile Val Tyr	Thr Ala Thr Asp Leu	Ser
	620	625	630
Gly Asn Gln Ala	Ser Cys Ile Phe His	Ile Lys Val Ile Asp	Ala
	635	640	645
Glu Pro Pro Val	Ile Asp Trp Cys Arg	Ser Pro Pro Pro Val	Gln
	650	655	660

Val Ser Glu Lys	Val His Ala Ala Ser	Trp Asp Glu Pro Gln Phe	
	665	670	675
Ser Asp Asn Ser	Gly Ala Glu Leu Val	Ile Thr Arg Ser His Thr	
	680	685	690
Gln Gly Asp Leu	Phe Pro Gln Gly Glu	Thr Ile Val Gln Tyr Thr	
	695	700	705
Ala Thr Asp Pro	Ser Gly Asn Asn Arg	Thr Cys Asp Ile His Ile	
	710	715	720
Val Ile Lys Gly	Ser Pro Cys Glu Ile	Pro Phe Thr Pro Val Asn	
	725	730	735
Gly Asp Phe Ile	Cys Thr Pro Asp Asn	Thr Gly Val Asn Cys Thr	
	740	745	750
Leu Thr Cys Leu	Glu Gly Tyr Asp Phe	Thr Glu Gly Ser Thr Asp	
	755	760	765
Lys Tyr Tyr Cys	Ala Tyr Glu Asp Gly	Val Trp Lys Pro Thr Tyr	
	770	775	780
Thr Thr Glu Trp	Pro Asp Cys Ala Lys	Lys Arg Phe Ala Asn His	
	785	790	795
Gly Phe Lys Ser	Phe Glu Met Phe Tyr	Lys Ala Ala Arg Cys Asp	
	800	805	810
Asp Thr Asp Leu	Met Lys Lys Phe Ser	Glu Ala Phe Glu Thr Thr	
	815	820	825
Leu Gly Lys Met	Val Pro Ser Phe Cys	Ser Asp Ala Glu Asp Ile	
	830	835	840
Asp Cys Arg Leu	Glu Glu Asn Leu Thr	Lys Lys Tyr Cys Leu Glu	
	845	850	855
Tyr Asn Tyr Asp	Tyr Glu Asn Gly Phe	Ala Ile Gly Pro Gly Gly	
	860	865	870
Trp Gly Ala Ala	Asn Arg Leu Asp Tyr	Ser Tyr Asp Asp Phe Leu	
	875	880	885
Asp Thr Val Gln	Glu Thr Ala Thr Ser	Ile Gly Asn Ala Lys Ser	
	890	895	900
Ser Arg Ile Lys	Arg Ser Ala Pro Leu	Ser Asp Tyr Lys Ile Lys	
	905	910	915
Leu Ile Phe Asn	Ile Thr Ala Ser Val	Pro Leu Pro Asp Glu Arg	
	920	925	930
Asn Asp Thr Leu	Glu Trp Glu Asn Gln	Gln Arg Leu Leu Gln Thr	
	935	940	945
Leu Glu Thr Ile	Thr Asn Lys Leu Lys	Arg Thr Leu Asn Lys Asp	
	950	955	960
Pro Met Tyr Ser	Phe Gln Leu Ala Ser	Glu Ile Leu Ile Ala Asp	
	965	970	975
Ser Asn Ser Leu	Glu Thr Lys Lys Ala	Ser Pro Phe Cys Arg Pro	
	980	985	990
Gly Ser Val Leu	Arg Gly Arg Met Cys	Val Asn Cys Pro Leu Gly	
	995	1000	1005
Thr Tyr Tyr Asn	Leu Glu His Phe Thr	Cys Glu Ser Cys Arg Ile	
	1010	1015	1020
Gly Ser Tyr Gln	Asp Glu Glu Gly Gln	Leu Glu Cys Lys Leu Cys	
	1025	1030	1035
Pro Ser Gly Met	Tyr Thr Glu Tyr Ile	His Ser Arg Asn Ile Ser	
	1040	1045	1050
Asp Cys Lys Ala	Gln Cys Lys Gln Gly	Thr Tyr Ser Tyr Ser Gly	
	1055	1060	1065
Leu Glu Thr Cys	Glu Ser Cys Pro Leu	Gly Thr Tyr Gln Pro Lys	
	1070	1075	1080
Phe Gly Ser Arg	Ser Cys Leu Ser Cys	Pro Glu Asn Thr Ser Thr	
	1085	1090	1095
Val Lys Arg Gly	Ala Val Asn Ile Ser	Ala Cys Gly Val Pro Cys	
	1100	1105	1110
Pro Glu Gly Lys	Phe Ser Arg Ser Gly	Leu Met Pro Cys His Pro	
	1115	1120	1125
Cys Pro Arg Asp	Tyr Tyr Gln Pro Asn	Ala Gly Lys Ala Phe Cys	

1130	1135	1140
Leu Ala Cys Pro Phe Tyr Gly Thr Thr Pro Phe Ala Gly Ser Arg		
1145	1150	1155
Ser Val Thr Glu Cys Ser Ser Phe Ser Ser Thr Phe Ser Ala Ala		
1160	1165	1170
Glu Glu Ser Val Val Pro Pro Ala Ser Leu Gly His Ile Lys Lys		
1175	1180	1185
Arg His Glu Ile Ser Ser Gln Val Phe His Glu Cys Phe Phe Asn		
1190	1195	1200
Pro Cys His Asn Ser Gly Thr Cys Gln Gln Leu Gly Arg Gly Tyr		
1205	1210	1215
Val Cys Leu Cys Pro Leu Gly Tyr Thr Gly Leu Lys Cys Glu Thr		
1220	1225	1230
Asp Ile Asp Glu Cys Ser Pro Leu Pro Cys Leu Asn Asn Gly Val		
1235	1240	1245
Cys Lys Asp Leu Val Gly Glu Phe Ile Cys Glu Cys Pro Ser Gly		
1250	1255	1260
Tyr Thr Gly Gln Arg Cys Glu Glu Asn Ile Asn Glu Cys Ser Ser		
1265	1270	1275
Ser Pro Cys Leu Asn Lys Gly Ile Cys Val Asp Gly Val Ala Gly		
1280	1285	1290
Tyr Arg Cys Thr Cys Val Lys Gly Phe Val Gly Leu His Cys Glu		
1295	1300	1305
Thr Glu Val Asn Glu Cys Gln Ser Asn Pro Cys Leu Asn Asn Ala		
1310	1315	1320
Val Cys Glu Asp Gln Val Gly Gly Phe Leu Cys Lys Cys Pro Pro		
1325	1330	1335
Gly Phe Leu Gly Thr Arg Cys Gly Lys Asn Val Asp Glu Cys Leu		
1340	1345	1350
Ser Gln Pro Cys Lys Asn Gly Ala Thr Cys Lys Asp Gly Ala Asn		
1355	1360	1365
Ser Phe Arg Cys Leu Cys Ala Ala Gly Phe Thr Gly Ser His Cys		
1370	1375	1380
Glu Leu Asn Ile Asn Glu Cys Gln Ser Asn Pro Cys Arg Asn Gln		
1385	1390	1395
Ala Thr Cys Val Asp Glu Leu Asn Ser Tyr Ser Cys Lys Cys Gln		
1400	1405	1410
Pro Gly Phe Ser Gly Lys Arg Cys Glu Thr Glu Gln Ser Thr Gly		
1415	1420	1425
Phe Asn Leu Asp Phe Glu Val Ser Gly Ile Tyr Gly Tyr Val Met		
1430	1435	1440
Leu Asp Gly Met Leu Pro Ser Leu His Ala Leu Thr Cys Thr Phe		
1445	1450	1455
Trp Met Lys Ser Ser Asp Asp Met Asn Tyr Gly Thr Pro Ile Ser		
1460	1465	1470
Tyr Ala Val Asp Asn Gly Ser Asp Asn Thr Leu Leu Leu Thr Asp		
1475	1480	1485
Tyr Asn Gly Trp Val Leu Tyr Val Asn Gly Arg Glu Lys Ile Thr		
1490	1495	1500
Asn Cys Pro Ser Val Asn Asp Gly Arg Trp His His Ile Ala Ile		
1505	1510	1515
Thr Trp Thr Ser Ala Asn Gly Ile Trp Lys Val Tyr Ile Asp Gly		
1520	1525	1530
Lys Leu Ser Asp Gly Gly Ala Gly Leu Ser Val Gly Leu Pro Ile		
1535	1540	1545
Pro Gly Gly Gly Ala Leu Val Leu Gly Gln Glu Gln Asp Lys Lys		
1550	1555	1560
Gly Glu Gly Phe Ser Pro Ala Glu Ser Phe Val Gly Ser Ile Ser		
1565	1570	1575
Gln Leu Asn Leu Trp Asp Tyr Val Leu Ser Pro Gln Gln Val Lys		
1580	1585	1590
Ser Leu Ala Thr Ser Cys Pro Glu Glu Leu Ser Lys Gly Asn Val		
1595	1600	1605

Leu Ala Trp Pro Asp Phe Leu Ser Gly Ile Val Gly Lys Val Lys  
 1610 1615 1620  
 Ile Asp Ser Lys Ser Ile Phe Cys Ser Asp Cys Pro Arg Leu Gly  
 1625 1630 1635  
 Gly Ser Val Pro His Leu Arg Thr Ala Ser Glu Asp Leu Lys Pro  
 1640 1645 1650  
 Gly Ser Lys Val Asn Leu Phe Cys Asp Pro Gly Phe Gln Leu Val  
 1655 1660 1665  
 Gly Asn Pro Val Gln Tyr Cys Leu Asn Gln Gly Gln Trp Thr Gln  
 1670 1675 1680  
 Pro Leu Pro His Cys Glu Arg Ile Ser Cys Gly Val Pro Pro Pro  
 1685 1690 1695  
 Leu Glu Asn Gly Phe His Ser Ala Asp Asp Phe Tyr Ala Gly Ser  
 1700 1705 1710  
 Thr Val Thr Tyr Gln Cys Asn Asn Gly Tyr Tyr Leu Leu Gly Asp  
 1715 1720 1725  
 Ser Arg Met Phe Cys Thr Asp Asn Gly Ser Trp Asn Gly Val Ser  
 1730 1735 1740  
 Pro Ser Cys Leu Asp Val Asp Glu Cys Ala Val Gly Ser Asp Cys  
 1745 1750 1755  
 Ser Glu His Ala Ser Cys Leu Asn Val Asp Gly Ser Tyr Ile Cys  
 1760 1765 1770  
 Ser Cys Val Pro Pro Tyr Thr Gly Asp Gly Lys Asn Cys Ala Glu  
 1775 1780 1785  
 Pro Ile Lys Cys Lys Ala Pro Gly Asn Pro Glu Asn Gly His Ser  
 1790 1795 1800  
 Ser Gly Glu Ile Tyr Thr Val Gly Ala Ala Val Thr Phe Ser Cys  
 1805 1810 1815  
 Gln Glu Gly Tyr Gln Leu Met Gly Val Thr Lys Ile Thr Cys Leu  
 1820 1825 1830  
 Glu Ser Gly Glu Trp Asn His Leu Ile Pro Tyr Cys Lys Ala Val  
 1835 1840 1845  
 Ser Cys Gly Lys Pro Ala Ile Pro Glu Asn Gly Cys Ile Glu Glu  
 1850 1855 1860  
 Leu Ala Phe Thr Phe Gly Ser Lys Val Thr Tyr Arg Cys Asn Lys  
 1865 1870 1875  
 Gly Tyr Thr Leu Ala Gly Asp Lys Glu Ser Ser Cys Leu Ala Asn  
 1880 1885 1890  
 Ser Ser Trp Ser His Ser Pro Pro Val Cys Glu Pro Val Lys Cys  
 1895 1900 1905  
 Ser Ser Pro Glu Asn Ile Asn Asn Gly Lys Tyr Ile Leu Ser Gly  
 1910 1915 1920  
 Leu Thr Tyr Leu Ser Thr Ala Ser Tyr Ser Cys Asp Thr Gly Tyr  
 1925 1930 1935  
 Ser Leu Gln Gly Pro Ser Ile Ile Glu Cys Thr Ala Ser Gly Ile  
 1940 1945 1950  
 Trp Asp Arg Ala Pro Pro Ala Cys His Leu Val Phe Cys Gly Glu  
 1955 1960 1965  
 Pro Pro Ala Ile Lys Asp Ala Val Ile Thr Gly Asn Asn Phe Thr  
 1970 1975 1980  
 Phe Arg Asn Thr Val Thr Tyr Thr Cys Lys Glu Gly Tyr Thr Leu  
 1985 1990 1995  
 Ala Gly Leu Asp Thr Ile Glu Cys Leu Ala Asp Gly Lys Trp Ser  
 2000 2005 2010  
 Arg Ser Asp Gln Gln Cys Leu Ala Val Ser Cys Asp Glu Pro Pro  
 2015 2020 2025  
 Ile Val Asp His Ala Ser Pro Glu Thr Ala His Arg Leu Phe Gly  
 2030 2035 2040  
 Asp Ile Ala Phe Tyr Tyr Cys Ser Asp Gly Tyr Ser Leu Ala Asp  
 2045 2050 2055  
 Asn Ser Gln Leu Leu Cys Asn Ala Gln Gly Lys Trp Val Pro Pro  
 2060 2065 2070  
 Glu Gly Gln Asp Met Pro Arg Cys Ile Ala His Phe Cys Glu Lys

2075	2080	2085
Pro Pro Ser Val Ser Tyr Ser Ile Leu Glu Ser Val Ser Lys Ala		
2090	2095	2100
Lys Phe Ala Ala Gly Ser Val Val Ser Phe Lys Cys Met Glu Gly		
2105	2110	2115
Phe Val Leu Asn Thr Ser Ala Lys Ile Glu Cys Met Arg Gly Gly		
2120	2125	2130
Gln Trp Asn Pro Ser Pro Met Ser Ile Gln Cys Ile Pro Val Arg		
2135	2140	2145
Cys Gly Glu Pro Pro Ser Ile Met Asn Gly Tyr Ala Ser Gly Ser		
2150	2155	2160
Asn Tyr Ser Phe Gly Ala Met Val Ala Tyr Ser Cys Asn Lys Gly		
2165	2170	2175
Phe Tyr Ile Lys Gly Glu Lys Lys Ser Thr Cys Glu Ala Thr Gly		
2180	2185	2190
Gln Trp Ser Ser Pro Ile Pro Thr Cys His Pro Val Ser Cys Gly		
2195	2200	2205
Glu Pro Pro Lys Val Glu Asn Gly Phe Leu Glu His Thr Thr Gly		
2210	2215	2220
Arg Ile Phe Glu Ser Glu Val Arg Tyr Gln Cys Asn Pro Gly Tyr		
2225	2230	2235
Lys Ser Val Gly Ser Pro Val Phe Val Cys Gln Ala Asn Arg His		
2240	2245	2250
Trp His Ser Glu Ser Pro Leu Met Cys Val Pro Leu Asp Cys Gly		
2255	2260	2265
Lys Pro Pro Pro Ile Gln Asn Gly Phe Met Lys Gly Glu Asn Phe		
2270	2275	2280
Glu Val Gly Ser Lys Val Gln Phe Phe Cys Asn Glu Gly Tyr Glu		
2285	2290	2295
Leu Val Gly Asp Ser Ser Trp Thr Cys Gln Lys Ser Gly Lys Trp		
2300	2305	2310
Asn Lys Lys Ser Asn Pro Lys Cys Met Pro Ala Lys Cys Pro Glu		
2315	2320	2325
Pro Pro Leu Leu Glu Asn Gln Leu Val Leu Lys Glu Leu Thr Thr		
2330	2335	2340
Glu Val Gly Val Val Thr Phe Ser Cys Lys Glu Gly His Val Leu		
2345	2350	2355
Gln Gly Pro Ser Val Leu Lys Cys Leu Pro Ser Gln Gln Trp Asn		
2360	2365	2370
Asp Ser Phe Pro Val Cys Lys Ile Val Leu Cys Thr Pro Thr Pro		
2375	2380	2385
Leu Ile Ser Phe Gly Val Pro Ile Pro Ser Ser Ala Leu His Phe		
2390	2395	2400
Gly Ser Thr Val Lys Tyr Ser Cys Val Gly Gly Phe Phe Leu Arg		
2405	2410	2415
Gly Asn Ser Thr Thr Leu Cys Gln Pro Asp Gly Thr Trp Ser Ser		
2420	2425	2430
Pro Leu Pro Glu Cys Val Pro Val Glu Cys Pro His Leu Arg Asn		
2435	2440	2445
Pro His Gly Ile Ile Asp Val Gln Gly Leu Ala Tyr Leu Ser Thr		
2450	2455	2460
Ala Leu Tyr Thr Cys Lys Pro Gly Phe Glu Leu Val Gly Asn Thr		
2465	2470	2475
Thr Thr Leu Cys Gly Glu Asn Gly His Trp Leu Gly Gly Lys Pro		
2480	2485	2490
Thr Cys Lys Ala Ile Glu Cys Leu Lys Pro Lys Glu Ile Leu Asn		
2495	2500	2505
Gly Lys Phe Ser Tyr Thr Asp Leu His Tyr Gly Gln Thr Val Thr		
2510	2515	2520
Tyr Ser Cys Asn Arg Gly Phe Arg Leu Glu Gly Pro Ser Ala Leu		
2525	2530	2535
Thr Cys Leu Glu Thr Gly Asp Trp Asp Val Asp Ala Pro Ser Cys		
2540	2545	2550



Asn Ala Ile His Cys Asp Ser Pro Gln Pro Ile Glu Asn Gly Phe  
 2555 2560 2565  
 Val Glu Gly Ala Asp Tyr Ser Tyr Gly Ala Ile Ile Ile Tyr Ser  
 2570 2575 2580  
 Cys Phe Pro Gly Phe Gln Val Ala Gly His Ala Met Gln Thr Cys  
 2585 2590 2595  
 Glu Glu Ser Gly Trp Ser Ser Ser Ile Pro Thr Cys Met Pro Ile  
 2600 2605 2610  
 Asp Cys Gly Leu Pro Pro His Ile Asp Phe Gly Asp Cys Thr Lys  
 2615 2620 2625  
 Leu Lys Asp Asp Gln Gly Tyr Phe Glu Gln Glu Asp Asp Met Met  
 2630 2635 2640  
 Glu Val Pro Tyr Val Thr Pro His Pro Pro Tyr His Leu Gly Ala  
 2645 2650 2655  
 Val Ala Lys Thr Trp Glu Asn Thr Lys Glu Ser Pro Ala Thr His  
 2660 2665 2670  
 Ser Ser Asn Phe Leu Tyr Gly Thr Met Val Ser Tyr Thr Cys Asn  
 2675 2680 2685  
 Pro Gly Tyr Glu Leu Leu Gly Asn Pro Val Leu Ile Cys Gln Glu  
 2690 2695 2700  
 Asp Gly Thr Trp Asn Gly Ser Ala Pro Ser Cys Ile Ser Ile Glu  
 2705 2710 2715  
 Cys Asp Leu Pro Thr Ala Pro Glu Asn Gly Phe Leu Arg Phe Thr  
 2720 2725 2730  
 Glu Thr Ser Met Gly Ser Ala Val Gln Tyr Ser Cys Lys Pro Gly  
 2735 2740 2745  
 His Ile Leu Ala Gly Ser Asp Leu Arg Leu Cys Leu Glu Asn Arg  
 2750 2755 2760  
 Lys Trp Ser Gly Ala Ser Pro Arg Cys Glu Ala Ile Ser Cys Lys  
 2765 2770 2775  
 Lys Pro Asn Pro Val Met Asn Gly Ser Ile Lys Gly Ser Asn Tyr  
 2780 2785 2790  
 Thr Tyr Leu Ser Thr Leu Tyr Tyr Glu Cys Asp Pro Gly Tyr Val  
 2795 2800 2805  
 Leu Asn Gly Thr Glu Arg Arg Thr Cys Gln Asp Asp Lys Asn Trp  
 2810 2815 2820  
 Asp Glu Asp Glu Pro Ile Cys Ile Pro Val Asp Cys Ser Ser Pro  
 2825 2830 2835  
 Pro Val Ser Ala Asn Gly Gln Val Arg Gly Asp Glu Tyr Thr Phe  
 2840 2845 2850  
 Gln Lys Glu Ile Glu Tyr Thr Cys Asn Glu Gly Phe Leu Leu Glu  
 2855 2860 2865  
 Gly Ala Arg Ser Arg Val Cys Leu Ala Asn Gly Ser Trp Ser Gly  
 2870 2875 2880  
 Ala Thr Pro Asp Cys Val Pro Val Arg Cys Ala Thr Pro Pro Gln  
 2885 2890 2895  
 Leu Ala Asn Gly Val Thr Glu Gly Leu Asp Tyr Gly Phe Met Lys  
 2900 2905 2910  
 Glu Val Thr Phe His Cys His Glu Gly Tyr Ile Leu His Gly Ala  
 2915 2920 2925  
 Pro Lys Leu Thr Cys Gln Ser Asp Gly Asn Trp Asp Ala Glu Ile  
 2930 2935 2940  
 Pro Leu Cys Lys Pro Val Asn Cys Gly Pro Pro Glu Asp Leu Ala  
 2945 2950 2955  
 His Gly Phe Pro Asn Gly Phe Ser Phe Ile His Gly Gly His Ile  
 2960 2965 2970  
 Gln Tyr Gln Cys Phe Pro Gly Tyr Lys Leu His Gly Asn Ser Ser  
 2975 2980 2985  
 Arg Arg Cys Leu Ser Asn Gly Ser Trp Ser Gly Ser Ser Pro Ser  
 2990 2995 3000  
 Cys Leu Pro Cys Arg Cys Ser Thr Pro Val Ile Glu Tyr Gly Thr  
 3005 3010 3015  
 Val Asn Gly Thr Asp Phe Asp Cys Gly Lys Ala Ala Arg Ile Gln

3020	3025	3030
Cys Phe Lys Gly Phe Lys Leu Leu Gly Leu Ser Glu Ile Thr Cys		
3035	3040	3045
Glu Ala Asp Gly Gln Trp Ser Ser Gly Phe Pro His Cys Glu His		
3050	3055	3060
Thr Ser Cys Gly Ser Leu Pro Met Ile Pro Asn Ala Phe Ile Ser		
3065	3070	3075
Glu Thr Ser Ser Trp Lys Glu Asn Val Ile Thr Tyr Ser Cys Arg		
3080	3085	3090
Ser Gly Tyr Val Ile Gln Gly Ser Ser Asp Leu Ile Cys Thr Glu		
3095	3100	3105
Lys Gly Val Trp Ser Gln Pro Tyr Pro Val Cys Glu Pro Leu Ser		
3110	3115	3120
Cys Gly Ser Pro Pro Ser Val Ala Asn Ala Val Ala Thr Gly Glu		
3125	3130	3135
Ala Pro Thr Tyr Glu Ser Glu Val Lys Leu Arg Cys Leu Glu Gly		
3140	3145	3150
Tyr Thr Met Asp Thr Asp Thr Asp Thr Phe Thr Cys Gln Lys Asp		
3155	3160	3165
Gly Arg Trp Phe Pro Glu Arg Ile Ser Cys Ser Pro Lys Lys Cys		
3170	3175	3180
Pro Leu Pro Glu Asn Ile Thr His Ile Leu Val His Gly Asp Asp		
3185	3190	3195
Phe Ser Val Asn Arg Gln Val Ser Val Ser Cys Ala Glu Gly Tyr		
3200	3205	3210
Thr Phe Glu Gly Val Asn Ile Ser Val Cys Gln Leu Asp Gly Thr		
3215	3220	3225
Trp Glu Pro Pro Phe Ser Asp Glu Ser Cys Ser Pro Val Ser Cys		
3230	3235	3240
Gly Lys Pro Glu Ser Pro Glu His Gly Phe Val Val Gly Ser Lys		
3245	3250	3255
Tyr Thr Phe Glu Ser Thr Ile Ile Tyr Gln Cys Glu Pro Gly Tyr		
3260	3265	3270
Glu Leu Glu Gly Asn Arg Glu Arg Val Cys Gln Glu Asn Arg Gln		
3275	3280	3285
Trp Ser Gly Gly Val Ala Ile Cys Lys Glu Thr Arg Cys Glu Thr		
3290	3295	3300
Pro Leu Glu Phe Leu Asn Gly Lys Ala Asp Ile Glu Asn Arg Thr		
3305	3310	3315
Thr Gly Pro Asn Val Val Tyr Ser Cys Asn Arg Gly Tyr Ser Leu		
3320	3325	3330
Glu Gly Pro Ser Glu Ala His Cys Thr Glu Asn Gly Thr Trp Ser		
3335	3340	3345
His Pro Val Pro Leu Cys Lys Pro Asn Pro Cys Pro Val Pro Phe		
3350	3355	3360
Val Ile Pro Glu Asn Ala Leu Leu Ser Glu Lys Glu Phe Tyr Val		
3365	3370	3375
Asp Gln Asn Val Ser Ile Lys Cys Arg Glu Gly Phe Leu Leu Gln		
3380	3385	3390
Gly His Gly Ile Ile Thr Cys Asn Pro Asp Glu Thr Trp Thr Gln		
3395	3400	3405
Thr Ser Ala Lys Cys Glu Lys Ile Ser Cys Gly Pro Pro Ala His		
3410	3415	3420
Val Glu Asn Ala Ile Ala Arg Gly Val His Tyr Gln Tyr Gly Asp		
3425	3430	3435
Met Ile Thr Tyr Ser Cys Tyr Ser Gly Tyr Met Leu Glu Gly Phe		
3440	3445	3450
Leu Arg Ser Val Cys Leu Glu Asn Gly Thr Trp Thr Ser Pro Pro		
3455	3460	3465
Ile Cys Arg Ala Val Cys Arg Phe Pro Cys Gln Asn Gly Gly Ile		
3470	3475	3480
Cys Gln Arg Pro Asn Ala Cys Ser Cys Pro Glu Gly Trp Met Gly		
3485	3490	3495



Arg	Leu	Cys	Glu	Glu	Pro	Ile	Cys	Ile	Leu	Pro	Cys	Leu	Asn	Gly	
			3500						3505					3510	
Gly	Arg	Cys	Val	Ala	Pro	Tyr	Gln	Cys	Asp	Cys	Pro	Pro	Gly	Trp	
			3515						3520					3525	
Thr	Gly	Ser	Arg	Cys	His	Thr	Ala	Val	Cys	Gln	Ser	Pro	Cys	Leu	
			3530						3535					3540	
Asn	Gly	Gly	Lys	Cys	Val	Arg	Pro	Asn	Arg	Cys	His	Cys	Leu	Ser	
			3545						3550					3555	
Ser	Trp	Thr	Gly	His	Asn	Cys	Ser	Arg	Lys	Arg	Arg	Thr	Gly	Phe	
			3560						3565					3570	

&lt;210&gt; 34

&lt;211&gt; 3696

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7722608CB1

&lt;400&gt; 34

```

atggggcaacg cggccggcag cgccgagcag cccgcggggc ccgcccgcgc gccccccaag 60
cagcccgcgc ctcccaagca gccgatgccc gcggccggag agctggaggc gcggttcaac 120
cgccccctga actgcatgaa cttgccccca gacaagggtcc agctgctgag ccagtatgac 180
aacgagaaga agtgggagct catctgtgat caggagcggg ttcaagtcaa gaatccccc 240
gcagcctaca tccagaagct gaagagctat gtggatactg gtggggtcag ccgaaaggta 300
gcagctgatt ggatgtccaa cctggggttt aagagggcag ttcaggagtc cagcagggtg 360
ctacggggagc tggagacctc cctgaggacc aaccacattg ggtgggtgca ggagtccctc 420
aatgaagaga accgtggcct ggatgtgctg ctccaggtacc tggcctttgc ccagtgtctt 480
gtcacgtatg acatggagag cacagacaac ggggcttcca actcagagaa aaacaagccc 540
ctggagcagt ctgtggaaga cctcagcaag ggtccaccct cctccgtgcc caaaagccgc 600
cacctgacca tcaagctgac cccagccacc agcaggaagg cctgcggaa ttcccgcctc 660
gtcagccaga aggacgacgt ccacgtctgt attatgtgcc tacgcgccat catgaactac 720
cagctctggct tcagccttgt catgaaccac ccagcctgtg tcaatgagat tgctctgagc 780
ctcaacaaca agaaccacag aaccaaggct ctggtgctgg agctgctggc ggccgtgtgc 840
ttggtgcggg gaggacatga catcatcctt gcagcctttg acaacttcaa ggaggtgtgt 900
ggggagcagc accgctttga aaagctgatg gaatatattc ggaatgagga cagcaacatc 960
gacttcatgg tggcctgcat gcagttcatc aacattgtgg tacattcggg ggagaacatg 1020
aacttccgtg tcttctgca atatgagttc acccatttgg gcttgacat gtacttggag 1080
aggcttcggc tcaccgagag tgacaagctg caggtgcaga tccaggcgta cctggacaat 1140
atthttgatg tggggcgct gctggaggac acagagacca agaacgctgt gctggagcac 1200
atggaggaac tgcaggagca agtggcgctg ctgacagagc ggcttcggga cgcggagaac 1260
gaatccatgg ccaagattgc agaactggaa aaacagctaa gccaggcgcg caaggagtgt 1320
gagaccctgc gggagcgctt cagcgaatcg accgccatgg gcgcctccag gcgtcccca 1380
gagcctgaga aagcgccctc cgctgccccg acgcggccct cggccctgga gctgaagggtg 1440
gaggagctgg aggagaagg gttaatccgt attctgcggg ggccggggga tgctgtctcc 1500
atcgagatcc tccccgtcgc tgtggcaact ccgagcggcg gtgatgctcc gactccgggg 1560
gtgccgaccg gctccccacg cccagatctc gcacctgcag cagagccggc tcccggagca 1620
gcgccaccgc cgccgccccc actgcccggc ctccccctcc cgcaggaagc cccgccctct 1680
gcgccccac agggcccgcc tctccctggc agcccggagc ccccgctgct gccgcccgtg 1740
ccggagacc tgcgccccc acccccgcca ccgccaccac ctccgggcac tgacgggccc 1800
gtgctccgc cgccgcccgc gccgcccggc cctcccggag gtccctctga tgccctagga 1860
agacgcgact cagaattggg cccaggagtg aaggccaaga agcccatcca gactaagttc 1920
cgaatgccac tcttgaactg ggtggcactg aaaccagacc agatcacagg cactgtcttc 1980
acagagctca atgatgagaa ggtgctgcag gagctagaca tgagtgattt tgaggaaacag 2040
ttcaagacca agtcccaagg cccagcctgt gacctcagcg ctctcaagag taaggcagcc 2100
cagaaggccc ccagcaaggc gacactcatt gaggccaacc gggccaagaa cttggccatc 2160
accctgcgga agggcaacct gggggccgag cgcacttgcc aagccattga ggcgtacgac 2220
ctgcaggctc tgggcctgga cttcctggag ctgctgatgc gcttcctgcc cacagagtat 2280
gagcgcagcc tcatcaccgg ctttgagcgg gagcagcggc caatggagga gctgtcagag 2340
gaggaccgct tcatgctatg cttcagccgc atcccgcgcc tgccggagcg catgaccaca 2400
ctcaccttcc tgggcaactt cccggacaca gccagctgct tcatgccgca actgaatgcc 2460

```

```

atcattgcag cctcaatgtc catcaagtcc tctgacaaac tccgccagat cctggagatt 2520
gtcctggcct ttggcaacta catgaacagt agcaagcgtg gggcagccta tggcttccgg 2580
ctccagagcc tggatgcgct gttggagatg aagtgcgactg atcgcaagca gacgctgctg 2640
cactacctgg tgaaggatcat tgctgagaag taccgcgaac tcacaggctt ccacagcgac 2700
ctgcacttcc tggacaaggc gggctcagtg tccctggaca gtgtcctggc ggacgtgcgc 2760
tccctgcagc gaggcctaga gttgacacag agagagtttg tgcggcagga tgactgcatg 2820
gtgctcaagg agttcctgag ggccaactcg cccaccatgg acaagctgct ggcagacagc 2880
aagacggctc aggaggcctt tgagtctgtg gtggagtact tccgagagaa cccaagacc 2940
acatcccag gcctgttctt ctccctcttt agcgcgttca ttaaggccta caagaaagct 3000
gagcaggagg tggaaacagt gaaaaaagaa gccgctgccc aggaggcagg cgctgatacc 3060
ccgggcaaaag gggagccccc agcaccgcaa aggcccgccg gccacagatg 3120
gacctcatct ctgagctgaa acggaggcag cagaaggagc cactcattta tgagagcgac 3180
cgtgatgggg ccattgaaga catcatcaca gatctgcgga accagcccta catccgcgca 3240
gacacaggcc gccgcagtgc ccgtcggcgt ccccgggcc cccactgca ggtcacctcc 3300
gacctctcgc tgtagccgt atttctgcag gtggattctg caggggtgtg gggccgtgga 3360
caggtgagg ctcaaggaa gtggctctca gctcggctgg ccgggcagcc cctcctccgc 3420
tgtggccgc ctcaaacggg ctggtgcatc ctctcttgg ccacagagg cagcatcgcc 3480
cgcccttcc cccaaatgct gcttgacgca cccaccctaa agccccctcc aaatagccat 3540
acttagcctc agcaggagcc tggcctgtaa cttataaagt gcacctcgcc ccgcaagcc 3600
ccagcccga ggaccgtcca tggaccttat tttatatga gattaataaa gatgtttgca 3660
aaaaaaaaa aaaaagcggc gctcgcgatc tagaac
3696

```

&lt;210&gt; 35

&lt;211&gt; 5585

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505869CB1

&lt;400&gt; 35

```

gaaagtgacc gtgtgcttcg gacggaccgg ggtggtcgtg ccgtgcgggg acggccacat 60
gaaagttttc agcctcatcc agcaggcggg gaccgcgtac cggaaggcca tcgccaagga 120
tccaaactac tggatacagg tgcctcgctt ggaacatgga gatggaggaa tactagacct 180
tgatgacatt ctttgtgatg tagcagacga taaagacaga ctggtagcag tgtttgatga 240
gcaggatcca catcacggag gtgatggcac cagtgccagt tccacgggta ccagagccc 300
agagatattt ggtagtgcgc ttggcaccaa caatgtctca gcctttcagc cttaccaagc 360
aacaagtga attgaggta cacccttcagt ccttcgagca aatatgcctc ttcattgttcg 420
acgcagtagt gaccagctc taattggcct ctccacttct gtcagtata gtaatttttc 480
ctctgaagag cttcaagga aaaatccac acgctggtca acaacagctg gcttctctca 540
gcagaacact gctgggagtc ctaaaacctg cgacaggaag aaagatgaaa actacagaag 600
cctcccgggg gatactagta actggtctaa ccaatttcag agagacaatg ctgcgtcgtc 660
tctgagtgcc agtcacccaa tgggtggcaa gtggctggag aaacaagaac aggatgagga 720
tgggacagaa gaggataaca gtcgtgttga acctgttggga catgctgaca cgggtttgga 780
gcatataccc aacttttctc tggatgatat ggtaaagctc gtagaagtcc ccaacgatgg 840
agggcctctg ggaatccatg tagtgccttt cagtgtcga ggcggcagaa ccctgggggt 900
attagtaaaa cgattggaga aaggtggtaa agctgaacat gaaaatcttt ttcgtgagaa 960
tgattgcatt gtcaggatta atgatggcga ccttcgaaat agaagatttg aacaagcaca 1020
acatatgttt cgccaagcca tgcgtacacc catcatttgg ttcattgtgg ttcctgcagc 1080
aaataaagag cagtatgaac aactatocca aagtgagaag aacaattact attcaagccg 1140
ttttagccct gacagccagt atattgacaa caggagtgtg aacagtgcag ggcttcacac 1200
ggtgcagaga gcaccccgac tgaaccaccc gcctgagcag atagactctc actcaagact 1260
acctcatagc gcacaccct cgggaaaacc accatccgct ccagcctcgg caccctcaga 1320
tgtattttagt acgactgtaa gcagtgggtta taacacccaa aaaataggca agaggttaa 1380
tatccagctt aagaaaggta cagaagggtt gggattcagc atcacttcca gagatgtaac 1440
aatagggtgc tcagctccaa tctatgtgaa aaacattctc ccccgggggg cggccattca 1500
ggatggccga ctttaaggcag gagacagact tatagaggta aatggagtag atttagtggg 1560
caaatcccaa gaggaagtgt tttcgctgtt gagaagcacc aagatggaag gaactgtgag 1620
ccttctgggtc tttcgccagg aagacgcctt ccacccaagg gaactgaatg cagagccaag 1680
ccagatgcag attccaaaag aaacgaaagc agaagatgag gatattgttc ttacacctga 1740
tggcaccagg gaatttctga catttgaagt cccacttaat gattcaggat ctgcaggcct 1800
tggtgtcagt gtcaaaggta accggtcaaa agagaaccac gcagatttgg gaatctttgt 1860

```

caagtccatt attaatggag gagcagcatc taaagatgga aggettcggg tgaatgatca 1920  
 actgatagca gtaaatggag aatccctgtt gggcaagaca aaccaagatg ccatggaaac 1980  
 cctaagaagg tctatgtcta ctgaaggcaa taaacgagga atgatccagc ttattgttgc 2040  
 aaggagaata agcaagtgcga atgagctgaa gtcacctggg agccccctg gacctgagct 2100  
 gcccatgtga acagcgttgg atgatagaga acgaagaatt tccattccc tctacagtgg 2160  
 gattgagggg cttgatgaat cgcacagcag aaatgctgcc ctcagtagga taatgggtga 2220  
 gtcaggtaaa taccagctgt cccctacagt gaatatgccc caagatgaca ctgtcattat 2280  
 agaagatgac aggttgccag tgcttctctc acatctctct gaccagtcct cttccagctc 2340  
 ccatgatgat gtgggggttg tgacggcaga tgctgggtact tgggccaagg ctgcaatcag 2400  
 tgattcagcc gactgctctt tgagtccaga tgttgatcca gttcttgctt ttcaacgaga 2460  
 aggattttgga cgtcagagta tgtcagaaaa acgcacaaaag caattttcag atgccagtca 2520  
 atttgatttc gttaaaacac gaaaatcaaa aagcatggat ttaggtatag ctgacgagac 2580  
 taaactcaat acagtggatg accagaaagc aggttctccc agcagagatg tgggtccttc 2640  
 cctgggtctg aagaagtcaa gctcgttggga gagtctgcag accgagttg ccgaggtgac 2700  
 tttgaatggg gatattcctt tccatcgtcc acggccgagg ataatcagag gcaggggatg 2760  
 caatgagagc ttcagagctg ccatcgacaa atcttatgat aaaccgagg tagatgatga 2820  
 tgatgaaggc atggagacct tggaagaaga cacagaagaa agttcaagat caggagagga 2880  
 gtcgtatcc acagccagtg atcagccttc ccactctctg gagagacaaa tgaatggaaa 2940  
 ccaagagaaa ggtgataaga ctgatagaaa aaaggataaa actggaaaag aaaagaagaa 3000  
 agatagagat aaggagaagg ataaaaatgaa agccaagaag ggaatgctga agggcttggg 3060  
 agacatgttc aggtttcaag ccaaaactcg agaatttagg gaacgacaag ctcgagagcg 3120  
 tgactatgct gaaattcaag attttcatcg gacatttggc tgtgatgatg agttaatgta 3180  
 tgggggagtt tcttcttatg aaggttccat ggctctcaac gctagacctc agagcccagc 3240  
 agaaggcat atgatggatg ctttgtatgc ccaagtcaag aagccgagg attccaaacc 3300  
 ctacactgtg gacagtaaca gatcaactcc tagcaatcat gatcggatc agcgtctgag 3360  
 gcaagaattt cagcaagcaa agcaagatga agatgtagaa gatcgtcggc ggacctatag 3420  
 ttttagcaa ccctggccga acgcacggcc ggcgacgagc agcggggcag actcgggtgc 3480  
 cgtggaggtg cagatgcagc ggcagcggca ggaggagcgc gagagctccc agcaggccca 3540  
 gcgccaatc agctctctgc ctgcggcaaa caggaaaaat gccagctcgg tctcccagga 3600  
 ctcttggggg cagaactact cccctgggga aggttccag agtgccaaag agaaccaccag 3660  
 gtactccagc taccaaggct ccaggaaagg ctacctggga ggacatggct tcaacgccag 3720  
 ggtcatgctg gaaactcagg agctccttcg ccaggaaacag aggcggaagg agcagcagat 3780  
 gaagaagcag cctccttccg aggggccag caactatgac tcgtataaga aagtcaggga 3840  
 cccagttac gccctccca aggggccctt ccggcaagat gtgccccct ccccttctca 3900  
 ggttgcgagg ctgaacagac ttcagactcc tgagaaaagg aggccttct attctgagc 3960  
 acgcaataaa cggatgcttc atgtcgcgca actatggagt atttctgtt ttggtatcag 4020  
 ttttgggagt ttttttaaaa cctcgatggt aaagaaatgg aaggccttaa tgtctttgcc actatgtctc 4140  
 tgcttctgt ttcatggaag gatttccac cctgtgacaa tcatctgttt gagggtgtca 4200  
 tatgtctgc gectctccac agtaccagga atctcggccc tactcatgag ttgtccgagg 4260  
 cttggttgta acatccctgc accacttgca gtgacaaatt cacctgaagt ggaggatgac 4320  
 gtgcggccct gtttctccct ctaagttctc ttagctatgg gatgacatct tagtctctgg 4380  
 tggaggaaaa gtgggcgaca tacacaaaaa attggggctt tctgggtact cacagcacag 4440  
 ccatttgcg tactttgtca tcaactgtgt tttctcttct ctttctcagc tctttgtgac 4500  
 gggagagtcg gtcatcctat tacagaagct aagccatagt caacattacc aataaagtga tgatcctggt 4560  
 tgggggtcct tttgtaactg ctttatgact gttcggctct gttcctgaca cgtgggttga gtcaccacag 4620  
 ctgcgtttat acatacgtt gttcggctct ggtcctgccc gggggaagct gggcctgccca 4680  
 ctctgtgtgg ggaacgtggg agacaggagt gggtcctgca taagaaagaa attctcctag gaagagctca 4800  
 ttggccctgt gtctatcatg aggggagagc attattttta ttagtttttg ttctgacagc ttgtcaggaa 4860  
 tggccagta catcctagta ataaaccaga cagtcatttt gatctgctct ctacggtttt 4920  
 gggcacagaa tgggacagag gcttgtctaa tgcatccaca cactgcatgt ctgactggcg 4980  
 tcaagtcaga ggcaattgat gcttgtctaa aacataaaaag acaaaggaaa agccgttaca 5040  
 atgccacgct cctaagtagt tctgccatga cgggtccac agcgggtggg gcaggaaagt cactctcggc 5100  
 catcacacag agaacatttt tgggtctcgg ggtctcggg ggtcccatc ttcttctcatt 5160  
 tcagtattag agtgtgtgtg taaacctggt ctgggggtgt aaagtgcctg tgaatcccca 5220  
 gttctgaaca tctgtattg aaagctcagc atacctgta ttactttta aaacagaaaa 5280  
 tgggttttag aaatgatgtg tccacacacg taccctgtgc cattgtattt ttgtgttgac 5340  
 tgggtctgtt tcttaatcac gtgcggcggt gtctaagtgg tgttaccagt gtacgcgag 5400  
 tgaccttga tgacagtggc tctttctcac agcctcccct gagctgtgag aaacagcttt 5520  
 ctctgtacat atgcaactcc taataaaagg catatttctt cctgttaaaa aaaaaaaaaa 5580  
 aaaaa

5585

<210> 36  
<211> 2667  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7502420CB1

<400> 36  
tgtttctgtt tagtttataa actgcgatcc agagtgtcag ctttcgtgcc ttgtgggact 60  
tgaagtccag aaggaagaag tgatcagggc tgccgagggg tgggaagcgtg cagccccctgc 120  
aggctctgaga cctacgaact acaactcccc gcaagcatcg gaaggcccgcg tcgtaattgc 180  
gggcaaccaa tcgtaccaga gcagggcaag cccccgccct ctggctcggg ggaaggcctc 240  
cgccaatggc agagctccca tcgcagagcc tgcgggttag gttgtgaggc ccggggccggg 300  
ggcggggagg agccaagggg gcgagcaagc tcgggtggctg ggtgggttgg ggcgttccgc 360  
gcgcccttca ttgaagcggc ggtggccggg ctggggcggc gtagtggaag gcgacggcgc 420  
ggctggaaaa tgccagtcca ttcccagagg gataagaagg agaccaacca tcacgatgag 480  
atggaggtgg actacgccga aaatgagggg agcagctccg aggacgagga cactgagagc 540  
tcgtcggctc cggaggatgg agatagctca gaaatggatg atgaagactg tgaaagaaga 600  
agaatggaaat gtttgatga aatgtccaat cttgaaaaac agtttaccga tctcaaagat 660  
caactttata aagaacgatt aagtcagggt gatgcaaaac tacaagaagt catagctgga 720  
aaagcaccag aataacttggg accgctggca actttacagg aaaatatgca aattcgtaca 780  
aaggtagcag gaatctatag agagctctgc ttagaatctg taaagaacaa atatgaatgt 840  
gaaattcaag cttctcgcca gcattgtgag agcgaaaagc tgttgctata tgatacagtc 900  
cagagtgaac tagaggagaa gataagaagg cttgaagagg ataggcacag cattgatatt 960  
acctcagagc tgtggaatga tgagcttcag tcaagaaaaa agagggaagga tcctttcagt 1020  
cctgacaaaa agaagccagt tgttgtttca ggtccatata tagtttata gctacaagat 1080  
cttgatattc ttgaagactg gacaacaatt aggaaggcaa tggctacatt ggggccacac 1140  
agagtgaaaa cggaaccacc tgtgaaactg gaaaaacatc tgcacagtgc tagatctgaa 1200  
gagggagac tatattatga tgggtaatgg tatatactg gacaaacaat atgtattgat 1260  
aaaaaagatg aatgtcctac aagtgtctga attacaacaa ttaaccatga tgaagtttg 1320  
ttaaagaggc ctgatggaag caaatctaa ctttacattt cacagctaca gaaaggaaaa 1380  
tattcaatta aacattcata atcatgattt aagtgttatt taaatttacc ttattagtgt 1440  
taccaaattg aagtgccatg agagtaaaaa aatgtattca ataacttaat attctcactg 1500  
aatcatgaga gaatgtgtat ttgtaggtag tactctaaat agatctcatt gatattgtat 1560  
taaaagaaac agtaataaaa attttatcac gatccttacg ttgatttgcc tcttaggtcc 1620  
gatgaccaat aggtattctg tatatggtag gggtttcttt ctaaacattt ttctttggtt 1680  
ttaaaaaaag ttatgcaaat ttgtcttacc tttagtaaac tatgactaca tttatctgca 1740  
atgtttctta taaatggttt gtactagtac attagtgtta ttgtaataa ggccgataga 1800  
catatatata tatgaggatg tatatatggc atcatcagct tatttagaac tgatggccat 1860  
accttacaat ctgtttttac ccaaaattaa gctattgggg ttgaaagcta aaaggagcac 1920  
ttttgtagaa tagcaacttt tcttttcttc tttcttgatt gtatgggtggg gtgggtgacct 2040  
atttttacaa attataccta atgagtaaaa ttagtgtaaa gtgataacat gcttctacct 2100  
gtattttctag tgacctttaa gcggcaggta ttatatactg gtatttatga tgcagtatat 2160  
aagtggtgaa caataactga cagtattgtg ctgctgtac atgtctgggc ttttgaaaca 2220  
gatttttagta agcattttcc agaggtaaaa ctgtgtcctt attctaattt tattcctagg 2280  
gcaaagtaga cagggattat ttccctgaat ctatttccaa attaatattt tttctttgg 2340  
tatttctaca ctttaaggcc atttggtgca atttagaaag tgttggcctc ccttccgcta 2400  
gccacattca aaattaaact ccaaaacctc aggaacagta caaagaattg aaacctcaa 2460  
tatggcagca cagccggctg tagtgtatat ttagggtgca ccaaatcagg tatctcgtgt 2520  
ggtcttgtgc actttaattt ctgttacaat gagttaagag gatgaggaag aaatctactt 2580  
atttaacact tactgcagaa atgtctgcat tattccgttt gttttcttat tattttacct 2640  
ctccaaaatc ttctgtgca gatcact 2667

<210> 37  
<211> 1559  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature

<223> Incyte ID No: 7506455CB1

<400> 37

```
ctgcctcctg agtgccggtg actgcggggag gcgacggagt gcttctgggg gtgcgagctg 60
gggaagttcg tgggtcacgga tgcgtgtggg gttgctgctc agtctgtaac ggcaggtggg 120
agacggacac tgggtggagg atggagtggc gaaagatgaa tgggagggct gattttcgag 180
agccgaatgc agaggttcca agaccaattc cccacatagg gcctgattac attccaacag 240
aggaagaaaag gagagtcttc gcagaatgca atgatgaaaag cttctgggtc agatctgtgc 300
ctttggctgc aacaagtatg ttgattactc aaggattaat tagtaaagga atactttcaa 360
gtcatcccaa atatggttcc atccctaaac ttatacttgc ttgtatcatg ggatactttg 420
ctggaaaact ttcttatgtg aaaacttgcc aagagaaaatt caagaaactt gaaaattccc 480
cccttgagga agctttacga tcaggacaag cagcagatc ttaccacctt gggcactatt 540
atcaaaagtc aaaatatgac tcaagtgtga gtggtcaatc atcttttgtg acatccccag 600
cagcagacaa catagaaatg cttcctcatt atgagccaat tccattcagt tcttctatga 660
atgaatctgc tcccactggg attactgatc atattgtcca agtcaaagta aacaagtatg 720
gagatacttg ggatgagtga aaaattacat cattggacat gaaggagttt caacatccag 780
cttcatctag gtggtcatga ttacctgcat gctttgagct cagcagcagt cttcataaac 840
acatttaaaa caagatcctg ggtttttgtg gtttgacttc tatggtgttt taaaaaaca 900
cagattttta gtgttaatat tgtgtaaag tactcacctt agggattcat ttgaatgatg 960
gtattatacc atgattgtat acagtttgtg aaattgttgc aaggggcaaag ataactctta 1020
aaaaaccgtc gagattacaa tgctctagaa tcagcatata agaaaataaa tgatatctgc 1080
atgttgaatt ggggtggatg gggggagcaa gcataatttt taagtgtgaa gctttgcac 1140
aagaaattat taaaaagctt ttttctcca gtattttctg tattatctta atgtttatgg 1200
caataaaaat gtaaaggaac atgcaacagc cctcatcttc cttgatttta tggttttatt 1260
gtttgttaatt tattgattta tttttaagct agttataatc atgtaggat aggaaataaa 1320
gtcatctata atatttctat aatatggcta taatatggct ataaatctat aatatggctg 1380
gaggcagtg gacgtggttc atgcctgtaa tcccagcact ttgggaggcc gaagtgggag 1440
gattgcatga gccaggagt ttgagaccag cctgggcaac aaagcgaacg aaaatcttta 1500
caaaaaaat taaaggaaaa aagtactcgg ccaaagctat cccctttagt gaggggttaa 1559
```

<210> 38

<211> 645

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7506018CB1

<400> 38

```
cgatgacttg gcgggaggag caggaaggaa accgctcccg agcacggcgg cggcgctcgtc 60
tcccgccagt gcagctgccg ctaccgccgc cctctgcccg ccggcccgtc tgtctacccc 120
cagcatgagc ggcctgcgcg tctacagcac gtcggtcacc ggctcccgcg aagactatga 180
gctcttcgtg gaggtgtgtg aacaaaacac gctgcaggag ttctgaagc tggcttgagt 240
caagcctgtc cagagttccc ctgctggact ccatcaccac actcccccca gccttcacct 300
ggccatgaag gaccttttga ccaactccct gtcatctcta acctaaccct agagtcacct 360
ccccaatgca ggccacttct cctccctcct ctctaaatgt agtcccctct cctccatcta 420
aaggcaacat tccttaccga ttagtctcag aaattgtctt aagcaacagc cccaaatgct 480
ggctgcccc agccaagcat tggggccgcc atcctgcctg gcactggctg atgggcacct 540
ctgttggttc catcagccag agctctgcc aaggccccgc agtcccctct ccaggaggac 600
cctagaggca attaatgat gtcctgttcc attggcaaaa aaaaa 645
```

<210> 39

<211> 1824

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 353017CB1

<400> 39

```
cctttattct gacaaagaga tgacaacccc agccaagcag attaagactg aaagtgaata 60
```

```

tgaagagact ttcaagccat catcagttgt gagtacctct tccactggct tgcggtttcaa 120
ccagccagag gaaggaaaca tcaactacaa gatatagtga tcacagtgtc acttccctcaa 180
ctctggcaca aattaatgct acttcttcag gaatgacaga agaatttctc cctccccccac 240
ctgacgtact tcaaaacttca gtagatgtga cagcattttc ccagtcacct gaactgcccc 300
gtcctcctag aagactacca gtccccaaag atgtatatc caagcaaaga aatttgtatg 360
aattaaaccg tttatataaa cacatccatc ctgagttaag aaaaaactta gaaaagatt 420
atatcagtga agtttctgag attgtttcta gtcaaatgaa ctcagggagt tcagtctcag 480
cagatgtgca acaagcccggt tatgttttga aaacacaaat gacagtctc aaaaagatct 540
gaactcagaa agagaatact tggaatggga tgaaattctg aaggaggagg tgcagtccat 600
tagatggatc tttgagaatc aaccattgga ttccatcaac aatggctctc ctgtatgaag 660
gtgatatttc caggggcatt gctgatcaag aaatcattgc tgggtggcgt gtgaaatata 720
ccacatggat gtttgaaacc caagccatcg acacacttgg ggctcattct tctgacactg 780
tagaaaatgc agagaaaatt cctgagctag ccagaggaga tgtctgcaca gctcgggtgga 840
tggttgaaac aaggccattg gactcaatga ataaaatgca tcaaagtcaa gaagaatcag 900
cggtaactat cagtaaggac ataactgggg gggatgtcaa gactgtgaga tacatgtttg 960
aaactcaaca tctagatcaa cttggacagc ttcatcaggt ggatgagggt catttactgc 1020
agcttaggtc tgagctcaaa gaaattaagg gaaatgttaa aagaagtata aaatgtttcg 1080
aaactcaacc attatatgtt attagagatg gttcgggcca aatgctggaa attaaaactg 1140
ttcacagaga agacgttgaa aagggagatg taagaacagc acggtggatg tttgaaacac 1200
agccgttgga cacaattaac ggggtgagta aggcaaagtg gttatttgaa acccaacctt 1260
tggaagaaaa tgtcaaaggt gaagaggtca cccagccatt agacattcta aaagaagttc 1320
tggaagaaaat caaagagtca gaagaggtca cccagccatt agacattcta aaagaagttc 1380
atgtctccag aaagtgttgg atgtttgaaa cccagccatt agacattcta aaagaagttc 1440
ctgatgcaga ttctctacaa cgtgaggaga taataggtgg tgatgtacaa actactaagc 1500
atctatttga aacacttcca attgaagcat taaaagacag tcctgatata ggaaagcttc 1560
aaaaaatca ctgcctctga agaagaaaaa ggggatgtta ggcatcaaaa atggattttt 1620
gaaacccaac ctctggaaga cattagaaaa gataaaaagg agtacacacg aacagtgaag 1680
cttgagaaa gttgacagag gagatgtgaa gaattacaca catatctttt gaatcaaaca 1740
atttcattaa atttgatgca tcacataaaa tagaggtggc aaggagttac agaaggtgtg 1800
taacagttaa ttaaatctct cttc

```

&lt;210&gt; 40

&lt;211&gt; 8043

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5137816CB1

&lt;400&gt; 40

```

gggggcgggtc ctgacggctc accgcaagaa gaaaagccac ccgttcgtcc cctcccccta 60
cctcccatcc tttgcccagg agctgccttg gcagtcacgc cccttccctc cgaggagctt 120
tctggctgcc taaactggta gaccccttga attactcctc catctccgct ctcttccgce 180
tctcttctc ttagttctct ccgctccccc ctcaactacc accacttcca gtcagtctcg 240
cctccggcta tccgctgctc caccctcttg ccgggtatcc tgcctgtccg ctgccaccaa 300
ggagagcccg gacggagcag cgaggagggg agcagccggg agttgggggt tccccctgc 360
ccatccctgg ccgctggccc gggaccgaag ccacttgagc gagcagagag tcgtcacctt 420
gtcttctttg ccttcaggga gctgctaaga aggacaaata agatagcaga gtgaaagagc 480
ttttgtctcc ttagaaggaa ggctgagaaa ctaaagagtg accagtgggt ggactcttgg 540
cggcgcttac cctgaacctg agtggttaggg gagagggaaat ctgccgttgc ggtttcttga 600
gctgctcata agcaagaatt ccctatcctg atcaaggctt taagcctaaa agaaagcaga 660
aatagggtcaa gggcagccca ggtgcccatt ttctctctc cttcacaagg caccaccagc 720
aatagagatg agaaattctg aagaacagcc aagtggaggg accacggttg cagcgtttgc 780
tacaagagca gcttcgctat ggcaatccta gtgagaatcg cagcctgctt gccatacacc 840
agcaagccac agggaatggc cctcctttcc ccagtggcag tgggaacccg ggccctcaga 900
gtgatgtgtt gactccccc aaccaccacc aacagcttgt ggctcatgct gctcgacaag 960
aaccacaggg gcaggaatc cagtcagaaa acctcatcat ggagaagcag ctgtctcctc 1020
gaatgcaaaa taatgaagaa ctcccgacct atgaagaagc caaggtccag tcccagtact 1080
ttcggggcca acagcatgcc agtgttggag ctgccttcta tgtcactgga gtcaccaacc 1140
agaagatgag gactgaggga cgcctatcag ttcagcgggt caatcctgga aagatgcacc 1200
aagatgagg actcagagac cttaagcaag ggcattgtccg ttctttagt gaacgactaa 1260
tgcagatgtc actggccacc agtggagtta aggccatcc acctgttacc agtgcctccc 1320

```

tctccccacc	acaacccaat	gacctctaca	agaatccac	aagttccagt	gaattctaca	1380
aggcccaagg	gccacttctt	aaccagcata	gcctgaaggg	catggaacac	cgaggccccc	1440
caccagaata	tcccttcaag	ggcatgccac	cccaatctgt	agtgtgcaag	ccccaagagc	1500
cagggcactt	ctatagttag	catcgctga	accagccagg	gagaacagag	gggcaactga	1560
tgaggtatca	gcacccccct	gagtatggag	cagccaggcc	agcgcaggac	atctcattgc	1620
cattgtcagc	caggaactcg	cagcctcaca	gccctacttc	ttctctgacc	tctggggggg	1680
ccttgccctt	gctacaatct	ccaccatcca	ctagattgtc	tcctgcccga	cacctcttgg	1740
tcccaaacca	gggagaccat	tcagctcacc	tgccataggcc	gcagcagcat	ttccttctcta	1800
atcaggctca	ccagggggat	cattaccgtc	tctcccaacc	tgccctgagt	cagcagcagc	1860
agcaacagca	gcagcagcac	catcatcacc	atcaccacca	acaacagcag	cagcagcagc	1920
cacagcagca	gccaggagaa	gcctattcag	ctatgcctcg	ggctcagcca	tcctctgctt	1980
cttatcagcc	agtgccagca	gacctttttg	ccattgtttc	cagagcccag	cagatgggtg	2040
agatcctctc	agacgagaac	cggaaacttga	ggcaagagtt	ggaaggatgc	tatgagaagg	2100
tggcaagact	gcagaaggtg	gagacagaaa	tccagcgctg	ctcggaggca	tatgagaacc	2160
tcgtgaagtc	atcctccaaa	agagaggccc	tagagaaagc	catgagaaac	aagctagagg	2220
gcgagattcg	gaggatgcat	gatttcaaca	gggatctgag	agagcgtcta	gagactgccca	2280
acaagcagct	tgacagagaag	gaatatgagg	ggtcagagga	caccagaaaa	acctctctgc	2340
agctcttttg	aaaaataaaa	gaaagccagc	gtgagaagga	gaagctggaa	gcggagctgg	2400
ccactgcccg	ttctaccaat	gaggacaaa	gacgacacat	cgaaatccga	gatcaggccc	2460
tgagtaatgc	ccaggccaag	gtggtaaagc	tggaagaaga	gctgaaaaag	aagcaagtgt	2520
acgttgacaa	ggtggagaag	atgcagcagg	ccctgttaca	gctccaggca	gcatgtgaaa	2580
aacgtgagca	gctagagcac	cgtctccgga	cacgactgga	gagggaactg	gaatccctga	2640
gaatccagca	gcgtcagggc	aactgtcagc	ccaccaacgt	ttcagaatac	aatgctgctg	2700
actgatgga	gctccttcgg	gagaagagg	agaggattct	ggctctggaa	gctgatatga	2760
caaagtggga	gcagaaatat	ttggaggaga	atgtgatgag	acattttgct	ctggatgctg	2820
ctgcaactgt	ggctgctcag	agggacacaa	cagtcacag	tcactctcct	aacaccagct	2880
atgacacagc	tctagaagct	cgcatccaga	aagaggagga	agaaatcttg	atggccaata	2940
agcgtttgct	tgacatggag	ggcaggatta	agaccctcca	tgcccagatt	attgagaagg	3000
atgccatgat	caaagtactc	cagcagcgtt	cccggaaagga	gccgagcaag	acagagcagc	3060
tgtcgtgcat	gcggccagcg	aagtctctga	tgtccatttc	caatgctgga	tcaggcttgc	3120
tctccactc	atccaccctg	actggctccc	ccatcatgga	agaaaagcga	gacgacaaga	3180
gctggaaggg	gagcctaggc	attctcctgg	gtggagacta	ccgtgctgaa	tatgtccctt	3240
ccacaccctc	gcctgtgccca	ccctcgactc	ccctgctctc	ggctcactcc	aagacaggca	3300
gccgagactg	cagtacccaa	actgaacgtg	ggacggaaatc	gaacaaaact	gcagctgttg	3360
tcccatctc	tgttcctgct	ccagttgctg	ctgccgccac	agcagcagcc	atcactgccca	3420
ctgctgcca	catcaccagc	atcatggtag	ctgctgctcc	agttgctggt	gctgctgctg	3480
ctgctccagc	tgctgctgct	gcccctgtct	cagccactgc	cgctgctact	gctgctgctg	3540
tttctccagc	tgctgctggt	cagattccag	ctgctgcctc	tgttgcctca	gctgctgccc	3600
ttgctccttc	tgctgctgct	gctgctgctg	ttcagggttg	tcagctgctg	ccggtccag	3660
ttccagctcc	ggctctggtt	ccggttccag	ctccagcagc	ggctcaggct	tctgctcctg	3720
ctcagactca	ggcaccactc	tcagctccgg	ctgtggctcc	aactccagct	ccaactccaa	3780
ctccagctgt	ggctcaggct	gaggttccctg	caagtccagc	taccggctct	ggaccacatc	3840
gtttgtctat	accaagtttg	acctgcaatc	cagacaaaac	agatgggcct	gtgttccact	3900
ccaatactct	ggaaagaaaa	actcccatte	agatcctggg	acaagagcct	gatgcagaga	3960
tggtggaata	tctcatctaa	acggccaaat	caagagctgc	agattatcag	caaaaatgct	4020
tttaatcatt	ttcccccttt	tattggttct	tgttttgagg	gtgaggacaa	gggttggtggg	4080
gaggggatgt	tttttaacag	gactttttat	tggaacaatg	tactacttga	gtaataccat	4140
gtgaacacca	gtctattttg	gtatgcttag	ggagtacctc	taaagacaga	ttaatcagaa	4200
tgtgctctaa	agcttattgt	ttgaatttat	acgaatactg	ggactgttaa	caggtggcta	4260
tacatcgacg	ttttcaatgt	gcttaaattt	gtttaaattt	tccatattct	agatcatttt	4320
ttattgaaga	gcacagtatg	tgtggaagac	agtgtataac	acgtagtgtg	gaagtgggaa	4380
gctagagaga	attgagtgtg	tgctgttttg	tatagttact	atcctgtgca	gcagctggag	4440
aaagcactca	cctcaggctt	acaaagggaa	tagtttcagg	agctatgtaa	gctggaaaaa	4500
aggtagggag	ttttgggggtg	cagaagggta	ctggagctaa	tttttcttcc	cactttccca	4560
gctaccctgc	cccagggaat	tgtgtttgtc	ttcatttcag	tggtgctttg	gaaatggatt	4620
cttttggttc	cctcctggag	gttcatacat	tcataatat	gctctggagt	aatttatgca	4680
tttgataaat	taataatttg	ctttcagatg	ctgggagagt	acattaactg	agtgatgcgc	4740
aacttcctct	ctcttaggga	attagaccat	cagaggcctt	gatggagagt	tgcatggggg	4800
gctatatgca	gacttccatg	gtttgtgtgt	agccatgaac	acagcttgct	tgcatttagt	4860
aagaccaatc	agcttagtgt	ttattttctc	tacagcacag	attcactggc	tgggtctcca	4920
gtctcaaatt	gccaatcatt	tgcaaatgta	ggaaggatct	ttgttgacag	gttgaatgct	4980
ttgaatttct	ggtgactact	ttgaaataac	ttgttttgtt	tgtcaaatte	taagcatatg	5040
tcttaaaagg	catttttgac	tatcacctcc	aagggaatag	cttgagaac	ccaaagtact	5100

```

atgctgcagt cgggggagag gtggattgca gcagtatcct caactacctc ttctcactgt 5160
cagtgcacac atcttgggaat acctttggga agcagcagga aatgtgcatg tgggtagaga 5220
tcaaaggagg caatggctcc aagccttgcc atagggctgc ctccaaggac acagaaggat 5280
gccagttgcc acaggctccct gccctgtgtc acctgtctgc cttcattaa ggtgagaaat 5340
ctgcagatag catcattaag atcagtttta aggggtatag ggagggtgag ggaagtgggg 5400
ggtgttaggt aagggttggg ggtagagggt ttgggatgtc ttagttagaa accagattaa 5460
tagaagagta ggctgatata attacatcat gagccatagt ggtgggaaag aacttttagca 5520
atatagccct acctctcat tttagtgtg aggaatctga gaactggaga ggttcagtga 5580
ctttttgaaa gtcatacaac acagctaacc attatgccaa tcaccatgct tattttggga 5640
aactctttat ctttttttaa ttccatttta tgaaaaggca tcttcattgg ccagggaata 5700
tgtatcttgt aaaatgtacc tgggtggagt agcttggtcca gtcttgacaa actactgaat 5760
ttctgtcttg cctctccttc agtgcctttt aaaagggtttt cccttttctg atctgcattt 5820
caacatagag tcacataaat gtccccctga gaaaccaatc ccacttcttt ctaggagatt 5880
gggtatctta gataatcttt tggggttcct ctgtgagtat aggaatggta tcttctctaa 5940
ttatcttcca aaggaattat tttgtgtgtg tgccctgtgtg tgtgtagaga cataaaggag 6000
ggtgatgtga ttttcagcta gtcctttcac attttcaata atgaggtaat catgttacat 6060
acacattagt cctcagttat aaagtgaatc tcagatagaa attaaaagtg cagttgtggt 6120
aagactcttt catactaccc tttagtcata aggagaaaaa aacactcaaa tagtagaagc 6180
agcaagtagc aaacttcagg agagctactt tctatccaaa taattttaaa aacacttttc 6240
acctactcct ttcattggtta taacacattg gcagactttt tgctggctct gggagccatg 6300
attttaatac cattctgcaa ggtgacaaat gtcatacatt ccacattgtg tggtagccat 6360
ctcttttagac tcatgtgttt tggggaaagg aagaagttct tggctgagta ctattttgaa 6420
ctttccagaa ccctctcaca ccagagacag ttcttctctg ttcagtttcc aatccccgat 6480
aatttgctaa aataacattg tacatccaag agagggaaga agagtatgtc agtatattat 6540
gcagaagata gatacagcct tttcagaaga tctccactag tttttgttcc aaaaattcaa 6600
gtttatggga gaaatctcaa ttagccacct tttcacagtt gtgtggatat aacatttggg 6660
ggatctttct ggactcctac ctatctgtgc attttaccgg cacctcagga aaggagggtg 6720
accaggttgt cttagcttgt actgcttggt gatctctgag gaccttctaa ttcagttgta 6780
ccccagttt ccattgtatag aaaaacttca ttagaacaaa ctttacttga tatgaaactc 6840
ctattaacag tttttttttg aaataaaaag tagcttgagc tttcttttaa aatcatgtat 6900
cttgattgtt gatttaatga aggaatttct tttaatgtc cttttgagct tcaaggtaat 6960
aggacagcag gaacctaaaa tatctgccat catctgccat aggaaagata cccagagacc 7020
catcatgttc tctttttgtt gttacactgt tgggtgggta taacaattgg aaaatgaaca 7080
aactgattga ttgtgcaaac tactttttat gacaagccta aacctcata atgcggcagc 7140
ttaaagtgtt tacatatgca ctaactttga tcaattatat tctcatatct gttagctaca 7200
cagtctccta ttatctcaat tgettattgt catatggaat atgttactta aaacgtgtgc 7260
attcttactg aaaatgtttt caaaggaagg tatcagctgt gggctaattg ccaccaattt 7320
cagcctgcca cgattcttgg aaatatgtct tccaagtgcc atccatcatc agtaggacaa 7380
gtgtcgggag tttgtttatt tttttccagt agcaacgatg ggttacatgg agccatgaaa 7440
cctccttctg gcctcccttg tgattaatgg catgtgtttg taaaatggat agctgggggt 7500
ggcagatggc tagagaagaa tcgcctttgg tttaaaatgt atgtgggtccc ctaatgattg 7560
tgaccccat ctgtaatcaa ctgagctagt tccaataaag ttaagcaggt ttaaattccac 7620
tttgtgcta tcttttcaact gacaataaag ttagctattt taaaatgcag taccgtgtgg 7680
aaaatgcttt ggtgtttacc cgacagaggc tttattttgt ctgtatcagt gattactttc 7740
aattcagtat gcagtgaat tgcttttcaa gggcagcgtg cagcagaatt tgcattttgt 7800
tgaagtatgg aatgtgatta tagattataa agtattaaga caacaccaca ggccgatgc 7860
tctaactcgg ctctttatat actcaggcag catatattaa aagctttgca tctttttatc 7920
ttctgggttc tagggataaa atgatggcca cccaaagcag agaatttatc ctttggttag 7980
cattttggaa ggccccctga tattcctttt gtacccttgt tctgatgcca attcatcagt 8040
ttt

```

<210> 41  
 <211> 1550  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7502151CB1

<220>  
 <221> unsure  
 <222> (1) ... (1550)



<223> a, t, c, g, or other

<400> 41

```

ctcgcacacg cgtcgccgac gctggtggcg acgctgtggg ggcaggtttt tagatcctca 60
gcggtttggac gagtttaatt agaatactaa gtaggtccgc ccagaagctt tcaccctgcc 120
ctctgaccct agggcccagc ccgaggaggc actcagggtt aggaggtctg ggcgagaagc 180
aactagggcc ctcatcactt cgcgcgcgaa tccccggcgc cgcacagcgg ggcagagcca 240
ggccagggcc gctcgcccaa cctggtccgc tgcctcttcg gccatggaag ctgccggcag 300
ccctgcggct acggagacag gcaagtatat cgcgtcaaca cagcgacctg acgggacctg 360
gcgcaagcag cggagggtga aagaaggata tgtgcccag gaggaggtcc cagtatatga 420
aaacaagtat gtgaagtttt tcaagagtaa accagagttg ccccagggc taagccctga 480
ggccactgct cctgtcacc cctccaggcc tgaagggtgt gaaccaggcc tctccaagac 540
agccaaacgt aacctgaagc gaaaggagaa gagggcgagc cagcaagaga aaggagaggc 600
agaggccttg agcaggactc ttgataagga gtccctggaa gagacagccc aactccccag 660
tgctccacag ggctctcggg cagccccac agctgcattc gaccagcctg actcagctgc 720
caccactgag aaagccaaga agataaagaa cctaaagaag aaactccggc aggtggaaga 780
gctgcagcag cggatccagg ctggggaagt cagccagccc agcaaagagc agctagaaaa 840
cctagcaagg aggaggcgctc tagaagagga gttagaggac ttggagttag gcctctgagg 900
cctttgggga atagggaatg gactgcagaa caaacctgtg ggctctctgg ggtctggggg 960
aatacgggca acagcagtca ggagggtac ccccatact ggcttacttc cactcctgc 1020
ggcccagctc tgcctccag agcctagcgt ctccctcaat ccttcccttt tcttcccaac 1080
tctactttt tggactttcc cctccatt cccagtgttc aaaatctcag tgactacccc 1140
aggtaccttt gctgctgatt tgggtgtctt gtttaaaaga aaatcagggtg ggtgggaatc 1200
tcttggagaa ctgaggctga gggtagaggg agtatccca agtcttggag tcttgggttc 1260
tgctcgcggt gtttatgggt tatttccctc tccatccctc atttttttt ttttttaaaa 1320
aaggcaaaat tgagaataaa cacaagtaga ctgtctgaa aaaaaaaaa aaaggggggg 1380
cgcgccagta tagggagctc gaaaccgggg gaataaattc ggagagggga acctggaggg 1440
ggggggagga aggagtccaa tatacaggta atgagaccgg gcagcctcag gggggcgcg 1500
ggtaccaaat acggcaaaaa gggagggnaga taagaggcgc aactggcaga 1550

```

<210> 42

<211> 2108

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7505983CB1

<400> 42

```

gtggcgagtg ccggccgaaa gctaggtccg gattgcacgt ggaggggcgc ccgaagggca 60
ctctcggaca ttaaccgcga ttctgtacca tggggcgcaa gttggaccct acgaaggaga 120
agcggggggcc aggcggaag gcccgaagc agaagggtgc cgagacagaa ctctcagat 180
tcttgctgc agtaagtgc gaaaattcca agaggctgtc tagtctgtct cgaaagaggg 240
cagccaagag gagattgggc tctgttgaag cccctaagac aaataagtct cctgaggcca 300
aaccattgcc tggaaagcta ccaaaaggga tctctgcagg agctgtccag acagctggta 360
agaagggacc ccagtcctta ttaatgctc ctgaggcaa gaagcggcca gcacctggca 420
gtgatgagga agaggaggag ttcttagaag ctaatgaggt gcctcggccc gtcacctcc 480
ggaccaatac cttgaaaacc cgacgccgag accttgaca ggctctaate aatcgtgggg 540
ttaacctgga tcccctgggc aagtgggtcaa agactggact agtgggtgat gattcttctg 600
tgccattgg tgctaccccc gagtacctgg ctgggcacta catgctgcag ggagcctcca 660
gcatgttgcc cgtcatggcc ttggcaccgc aggaacatga gcggatcctg gacatgtgtt 720
gtgcccctgg aggaaagacc agctacatgg ccagctgat gaagaacacg ggtgtgatcc 780
ttgccaatga cgccaatgct gagegggtca agagtgttgt gggcaacttg catcggttg 840
gagtcaccaa caccattatc agccactatg atgggcgcca gttccccaag gtgggtgggg 900
gctttgaccg agtactgtc gatgctccct gcagtgccac tggggtcac tccaaggatc 960
cagccgtgaa gactaacaag gatgagaagg acatcctgcg ctgtgtctac ctccagaagg 1020
agttgtcctt gagtgtctat gactctgtca atgcgacctc caagacagga ggctacctgg 1080
tttactgcac ctgttctatc acagtagaag agaattagtg ggtggtagac tatgctctga 1140
aaaagaggaa tgtgcgactg gtgcccacgg gcctagactt tggccaggaa ggttttacc 1200
gctttcgaga aaggcgcttc caccacagtc tgcgttctac ccgacgcttc taccctcata 1260
cccacaatat ggaagggttc ttcatgcca agttcaagaa attttccaat tctatccctc 1320
agtcccagac aggaaattct gaaacagcca cactacaaa ttagacttg cctcaggtca 1380

```

```

tccccaagtc tgagaacagc agccagccag ccaagaaagc caagggggct gcaaagacaa 1440
agcagcagct gcagaaacag caacatccca agaaggcctc cttccagaag ctgaatggca 1500
tctccaaagg ggcagactca gaattgtcca ctgtaccttc tgtcaciaag acccaagctt 1560
cctccagctt ccaggatagc agtcagccag ctggaaaagc cgaagggatc agggagccaa 1620
aggtgactgg gaagctaaag caacgatcac cttaaattaca gtctctcaag aaagttgctt 1680
tcctcaggca gaatgcccc cccaagggca cagacacaca aacaccggct gtgttatccc 1740
catccaagac tcaggccacc ctgaaacctc aggaccatca tcagccccctt ggaagggcca 1800
aggggggttg gaagcagcag ttgccagagc agccttttga gaaagctgcc ttccagaaac 1860
agaatgatac ccccaagggg cctcagcctc ccactgtgtc tcccatccgt tccagccgcc 1920
ccccaccagc aaagaggaag aaatctcagt ccaggggcaa cagccagctg ctgtctatctt 1980
agatgggttg aaactagacg ggtgggtcac tgccattgtc accaggttgg aactcttgcc 2040
tctgtgagga tgctttctct actgtgcata cccatgaaat ttaatacaca ttttaaaacc 2100
tctgaaaaa                                     2108

```

<210> 43  
 <211> 524  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7505986CB1

```

<400> 43
gtggcgagtg ccggccgaaa gctaggtccg gattgcacgt ggagggccgc ccgaagggca 60
ctctcggaca ttaaccgca ttctgtacca tggggcgcaa gttggaccct acgaaggaga 120
agcggggggcc aggccgaaa gcccggaaag agaagggtgc cgagacagaa ctctgcagat 180
tcttgcttgc agtaagtgc gaaaattcca agaggctgtc tagtcgtgct cgaaagaggg 240
cagccaagag gagattgggc tctgttgaag cccctaagac aaataagtct cctgaggcca 300
aaccattgcc tggaaagcta ccaaaagggg tctctgcagg agctgtccag acagctggta 360
agaaggggacc ccagtcacct tttaatgctc ctcgaggcaa gaagcgccca ttgtcaccag 420
gttggaaactc ttgctctgtg gaggatgcct tctctactgt gcatacccat gaaatttaac 480
acacattttta aaacctctga aaaaaaaaaa aaaaaaaaaa aagg                                     524

```

<210> 44  
 <211> 1987  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3231075CB1

```

<400> 44
ttctgcttat catggttacc aacaaaatga ctgctgcctt tagaaaccct agtgggaaac 60
aggtggcgac agacaaagtt gcagaaaagc tgagctctac tctctcatgg gtgaagaaca 120
cagtatcgca tacagtcagt cagatggcca gtcagggtggc aagtccatct acttcattac 180
ataccacatc ctcatctacc acactatcaa caccagccct ttaccatct tccccatcac 240
agttgagtc agacgtctta gaactcctgg ctaaaactgga agaacagaat atattgttag 300
aaacggatag taagtcttta agatctgtaa atgggtcaag aagaaacagt ggctcttctc 360
ttgtgtcgag ttcatcagcc tctagcaacc tcagtcacct tgaagaagat tcttggattc 420
tttggggaag aattgttaat gaatgggaag atgtacgcaa aaagaaggaa aagcaagtta 480
aggaacttgt tcataaaggg ataccccatc actttagagc aatagtttgg caacttttat 540
gcagtgcaca aagtatgcca attaaggatc agtattcaga actcctgaaa atgacctcgc 600
cttgtgaaaa attgatccga agggacattg ctagaactta ccctgaacac aactttttta 660
aggaaaaaga tagccttggg caggaggttt tatttaattg aatgaaggct tactctttag 720
tagatcgtga ggttggttac tgtcaaggaa gtgcttttat agttggattg ttgcttatgc 780
agatgccaga agaagaagct ttctgtgtat ttgttaaat aatgcaagat tatagacttc 840
gtgaactttt taaaccaagt atggcagaat tgggcctttg tatgtaccag tttgaatgta 900
tgatacagga gcatcttcca gagctctttg tacattttca atctcagagt tttcatacct 960
caatgtatgc atcatcctgg tttctgacta tctttcttac aacttttcca ctaccagttg 1020
caacaaggat atttgatata tttatgtctg agggtttaga aatagtgttt cgtgtaggat 1080
tagcacttct tcagatgaat caggcagaac tgatgcaact tgacatggaa gggatgttac 1140

```

```

agcactttca aaaggtcatt ccacatcagt ttgatgggtg cccagacaag ctaatccaag 1200
cagcttacca agtcaaatac aattcaaaaa aaatgaaaaa gcttgaaaaa gaatacacta 1260
caataaaaaa gaaagaaatg gaagagcaag ttgaaattaa aagggttacgc acagaaaaata 1320
gacttttaaa acagcgcata gagacattag aaaaacataa atgcagttcc aactacaacg 1380
aagattttgt gctacagcta gagaaggaat tgggtccaagc ccgactgagt gaagctgagt 1440
ctcagtgtgc attaaaagag atgcaggata aagtcttgga tatagagaag aggaataact 1500
cccttcctga tgagaataat attgcaaggc ttcaggaaga actcattgct gtgaaactta 1560
gagaagcaga agccattatg ggtttgaaag aacttagaca gcaagtcaag gatttagagg 1620
aacactggca gcgccactta gctcgtacta ctgggagatg gaaagaccca cccaagaaaa 1680
atgctatgaa tgagttacag gatgaactga tgaccattcg acttagagaa gctgaaacac 1740
aagcagaaat aagagaaata aaacaaagga tgatggaaat ggaaacacag ttggttgaat 1800
ccgcagatgt ggagcctgca tatatgaagg gctgaccgta tactctctgc tcttccaga 1860
gtatctggca tgtgatgagc tcgccacaaa catttagtga ttcgagtaag gctctggtg 1920
atttacagtt tctactttcc tgacactctg ttgctgggtt gaataaaaaa tcttctctga 1980
aaaaaaa
1987

```

&lt;210&gt; 45

&lt;211&gt; 2857

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7503516CB1

&lt;400&gt; 45

```

ctgccgggtg tgccgggtgt ccagcgaacc cctttcccaa accttcgggg agaagggagg 60
tgaggaggag caaagaaact acaggcaggg agctggaagg gggggtgggg ggggcaggag 120
acaagaaatc aagacaccag gcagcaggac acacacacac tcacatacac tcacacacat 180
agagaccaac agatagacag ctacctaaag cctgaaagac tgacagcaac acagaaaaaa 240
agaaacaggc agaaagagag acaaagacag aaatagaaac agactaacac acagagtcaa 300
aaatacagag acagaaagac agggagaaaag agaaacagaa aattagacac caaagacata 360
cgaacaggga ggaaggccga ctgaaagaaa gacggagaag aggagagaga agccagggcc 420
gagcgtgccg gcaggcggat ggagggcgcc ctggtggagg aggagacgta gtggcctggg 480
ctgagctggg tgggcccggg gaagcgggtg cctcagagtg ggggtggggg catgggaggg 540
gcaggcattc tgctgctgct gctggctggg gcgggggttg tgggtggcctg gagaccccca 600
aagggaagt gtccctcgcc ctgctcctgc tctaaagaca gcgcctgtg tgagggctcc 660
ccggacctgc ccgtcagctt ctctccgacc ctgctgtcac tctcactcgt caggacggga 720
gtcaccacgc tgaaggccgg cagcttcctg agaattccgt ctctgcacct gctcctcttc 780
acctccaact ccttctccgt gattgaggac gatgcatttg cgggcctgtc ccacctgcag 840
tacctgtgag tgaggccagg ggatggaggc agcttcacg aggacaatga gattggctcc 900
atctetaaga atgccctcag aggacttcgc tcgcttacac acctaacgct ggccaataac 960
catctggaga cctcccccag attcctgttc cgaggccttg acaccttac tcacgtggac 1020
ctccgcgggg acccggtcca gtgtgactgc cgcgtcctct ggctcctgca gtggatgcc 1080
accgtgaatg ccagcgtggg gaccggcgcc tgtcggggcc ccgcctccct gagccacatg 1140
cagctccacc acctcgacc caagacttcc aagtgcagag ccatagagct gtcctggttc 1200
cagacgggtg gggagtccgc actgagcgtg gagcccttct cctaccaagg ggagcctcac 1260
attgtgctgg cacagccctt cgcggccgc tgctgtatc tctcctggga ctacagcctg 1320
cagcgtctcc ggcccagga agagctgccc gccgcctgt tgggtgtcctg caagccactg 1380
gtgctgggcc ccagcctctt cgtgctggct gccgcctgt ggggggggtc acagctgtgg 1440
gcccgcccca gtcccgccct gcgcctggcc ccaacgcaga ccctggcccc gcggcggtg 1500
ctgcggccca atgacgccga gctcctgtgg ctggaagggc aacctgctt cgtgggtggc 1560
gatgcctcca aggcgggcag caccacgtg ctgtgccgcg acgggcccg cttttaccg 1620
caccagacc tgacgcctg gcaccggag acggacgtg aggcctgga gctggacggc 1680
cgccccacc tgctgctggc ctcggcttcc cagcgcccg tgctcttcca ctggaccggt 1740
ggcgcttcg agagacgcac ggacatcccc gaggcgagg atgtctatgc cacacgcac 1800
ttccaggctg gtggggacgt gttcctgtgc ctcacacgt acattgggga ctccatggtc 1860
atgcgctggg acggtcccat gttcgtctg ctgcagcaac ttccctcgcg cgggtgccac 1920
gtcttcacgc cactgctcat cgcaggggac cagctggcca tcttaggcag cgacttcgcc 1980
ttcagccagg tctccgcct tgagcctgac aagggggtc ctggagccac tcgaggagct 2040
ggggcctccg gccctgggtg cccccgtgc ctttgcacac atcactatgg ccggcagacg 2100
cttctctttt gctgcttgct ttaagggccc cacacagatc taccagcatc acgagatga 2160
cctcagtgcc tgagaccacc aacgggactc tgggcatggc tggggccctt ggacggcccc 2220

```

```

ttggctggct cctggcccta cttgggggtga tggcccgcct gtgagctgct gaccgtgggc 2280
cacgttcacg agccacacgt ctaggcctta agcccacttc ttaaaggatc tgcacccatg 2340
gggggacata gaggggtcca gcctactgga cctcccgcgc tgccttcag gtctaagca 2400
acatctggac atctcggtcg gggcagattc cccatcaaag gcaacctggg ggggtgcaggc 2460
tgggtgactg catggagtcg aggtggaggt cacgtgcaga gctccatgag gtcccggagc 2520
ccccctctca acccgctcct accccgctgg tccctctgaa agcaccgagg ggttgatggc 2580
attcctttcc cattctaaac tccacgcagc caatcgcttt cctggaggag cgccaacctc 2640
gaagacaact cagcactggg cgggttcccc tctgctgctt gtgcgcatgc gtgcggccgc 2700
ccagcaagct gtgccttttg cgcctcttgc gcgtgcgcgg ggcacaggga gggacagcgg 2760
gaggtgtcgc ggggagcctg gtgccttct agatgggcga cggaggcaag gacaagacga 2820
gtccgagctc aataaatgcg ctgtgcaccc ccaaaaaa 2857

```

&lt;210&gt; 46

&lt;211&gt; 2375

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506179CB1

&lt;400&gt; 46

```

cggcacgagg cacattcctg gggagtccct agctgccage atctgattag aaccatatct 60
ctcgccggga gtggccgcgc ggctccgaag ctcccggccg ggggtattt aagcgaggcc 120
cgccgcaccc gctgcgctgt agcctggagg ctccgggcgc ggggaagtc tgcctccttc 180
acggaggcaa tagctagccg gtgtctgtgg gaggttatgt ttatttgaga gttctccatc 240
gggatcgccct ggtgtcacca agtgtccact ggtactgagg tttgtgcct gccttcttgc 300
catgtctaac gaagttagaaa caagtgcac caatggtcag ccgaccaac aggcgcgacc 360
aaaagcacc tcaagaagg aaaaaagaa aggcctgaa aagacagatg aatatctctt 420
agcaagggtc aaaggcgatg gtgtaaaata taaggccaag ctgattggca ttgatgatgt 480
gccagatgca agaggggata aaatgagcca agactctatg atgaaactaa agggaatggc 540
ggcagctggg cgggtctcagg gacaacacaa acaaggatc tgggtcaaca tttccctttc 600
tgggataaaa ataattgatg agaaaactgg ggtaatagag catgaacatc cagtaataaa 660
gatttctttc attgcccgtg atgtgacaga caaccgggca tttggttacg tgtgtggagg 720
agaaggccag catcagtttt ttgccataaa aaccgggcaa caggctgaac cattagttgt 780
tgatcttaaa gacctttttc aagttatcta taatgtaaag aaaaaggaag aagaaaagaa 840
aaagatagag gaagccagca aagcagttga gaatgggagt gaggccctaa tgattctaga 900
tgaccaaaact aacaaaactg aatcgggtgt tgaccagatg gatttgtttg gggacatgtc 960
tacacctcct gacctaaata gtccaacatc ttcagccaat gacttgcttg catcagacat 1020
ctttgtcctt cccgtctcag aaccttcagg ccaggcgta cccacaggac aacctacagc 1080
cctgcagccc aacctcttgg atctcttcaa acaagtgtc cctgccccag tggggccctt 1140
ggtgggtcta ggtggtgtaa ctgtcacact ccctcaggca ggaccatgga acacagcatc 1200
tttggctctt aatcagtcct ctccaatggc tccgggagcc atgatgggtg gtcaaccttc 1260
aggttttagt cagcccgta tttttggtac aagtcagct gtttcagggtt ggaaccagcc 1320
ttcacctttt gcagcctcaa ctcccctcc agtgccgtt gtctggggcc ctctgcac 1380
tgtggcacc aatgcttgg caacaacaag ccctttgggg aatccttttc agagcaatat 1440
ttttccagct cctgctgtgt ccaactcagc cccatccatg cactcctctc tcttggtcac 1500
tcctcctcag ccacctccca gagctggccc tccaaggac atctccagtg atgcttcac 1560
tgcttagaca ccacttgggg ataaaagatg caaggatgtg aaagaaatgt ttaaggattt 1620
ccaactgcgg cagccacctg ctgtgcccgc gcggaaggga gagcagactt ctctgggac 1680
tttgagtgc tttgacagtt atttcaacag caaggttggc attcctcagg agaatgcaga 1740
ccatgatgac tttgatgcta atcaactatt gaacaagatc aatgaaccac caaagccagc 1800
tcccagacaa gtttccctgc cagttaccaa atctactgac aatgcatttg agaacctttt 1860
ctttaaagat tcttttgggt catcacaagc ctctgtggct tcttctcaac ctgtatcttc 1920
tgagatgtat agggatccat ttggaaatcc ttttgctaa attctgaact tgggtctgcag 1980
accatccaga ggaataaaaa ggttggcctt agtagtcaa aacaaagctg atagccagac 2040
acgttctgat tctgcccctt gttccagctt tgacgtatta tctgttgcct tatttctcat 2100
tgctctttct acttgtaaaa tgcttttcac tttctgtcta ggttaaagct aaactgaatc 2160
tatggcttta aataaattaa gatcctaacc tctctagctt aagtgtaaat gaagtacagt 2220
agtttcccta ctgaacctcg cctcttgtgt ccttggaaac ttctagaaca cctgccttct 2280
acctctgggt tgggagatgc agccaccaca tcccttcata tcatactgtt ttgaataaat 2340
tttcaaatcc ttattgttaa aaaaaaaaaa aaaaaa 2375

```

<210> 47  
 <211> 486  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1938744CB1

<400> 47  
 gaagaggcag atggttaggaa atggagacgg agaaatggca ctaatggccc agccacctga 60  
 gatgggatac ctgggagggga catcttttca agtagaggct gtgttggggc atggtgagag 120  
 agggctcttag caggtaacct tcctttcctc tccagactga ggaatcagag ttctgattgt 180  
 ggagtgcctc tctctaggac ggggctgcag cataggagtc tcagctgctt acatccaggt 240  
 ccaggattat gtctgctaac agacgctggg ggggtaccacc tgacgacgaa gactgtgtgt 300  
 ctgagaagct cctgaggaag actcgggaat ctccactggg gcctataggt gctgtgtaca 360  
 caatgtacag cgattacgtc aagaggatgg cacaggatgc tggagagaag taggactcct 420  
 ataggagccg gggctgtcca actcccctaa ctcaatccct ggtacattcc taataaagca 480  
 gttttg 486

<210> 48  
 <211> 1527  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5557436CB1

<400> 48  
 ggcgctgcgg gacggctagc ggccctgcgt gggagtccct gacagctgga ttctagaagt 60  
 agaactatga gctcaccttt ggccctccctt agcaagaccc gaaaagtgcc cctgccctcg 120  
 gagcctatga atcctgggag gcgaggaatc cgcacttatg gagatgtccc tgcaccccat 180  
 gactcggagc tgatggcctt catgacgagg aagttgtggg acctggagca gcaggagaag 240  
 gccagactg atgagatact gtccaaggat cagaagatag cggccctaga ggacctggtg 300  
 cagaccctcc ggccacaccc agccgaggca accctgcagc ggcaggagga actggagacg 360  
 atgtgtgtgc agctgcagcg gcaggtcagg gagatggagc ggttcctcag tgactatggc 420  
 ctgcagtggg tgggcgagcc catggaccag gaggactcag agagcaagac agtctcagag 480  
 catggcgaga gggactggat gacagccaag aagttctgga agccagggga ctcatggcg 540  
 cccctgagg tggactttga caggctgctg gccagcctgc aggatcttag tgagctggtg 600  
 gtagaggggtg acacccaagt gacaccagt cccggcgggg cagggtgcg taccctcgag 660  
 cccatcccg cgaagctcta ccggaatggc atcatgatgt tcgacgggcc ctccagccc 720  
 ttctacgac cctccacaca gcgctgcctc cgagacatat tggatggcct ctttccctca 780  
 gagctccagc gactgtatcc caatgggggtc cctttaagg tgagtgaact gcgcaatcag 840  
 gtctacctgg aggatggact ggacccttc ccaggcgagg gccgtgtggt gggcaggcag 900  
 cggatgcaca aggccttgga cagggtggag gagcaccag gctccaggat gactgctgag 960  
 aaatttctga acaggctccc caagtttgtg atccggcaag gcgagggtgat tgacatccgg 1020  
 ggcccatca gggacacctt gcagaactgc tgcccattgc ctgcccggat ccaggagatt 1080  
 gtggtggaga cgccacctt ggccgctgag cgagagagga gccaggagtc acccaacaca 1140  
 ccggcacccc cgctctccat gctgcgcatc aagtctgaga atggggaaca ggccttcta 1200  
 ctgatgatgc agcctgacaa caccattggg gacgtgcgag ctctgctagc gcaggccagg 1260  
 gtcatggatg cctctgcctt tgagatcttc agcacattcc cgccaccct ctaccaggac 1320  
 gatacactca cgctgcagge tgcaggcctt gtgcccgaag cagcactgct gctgcgggca 1380  
 cgccgagccc cgaagtccag cctgaaatc agtctgtgtc cctgtcccg tcccggctcc 1440  
 ggccccagtc ccggtcccg tcccgtctcc agtcctgtc ccggaccag tcccagcccc 1500  
 caataaagca cccgccccct ctaaaaa 1527

<210> 49  
 <211> 1422  
 <212> DNA  
 <213> Homo sapiens

<220>

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506178CB1

&lt;400&gt; 49

```

ggcgaagacg gagttgtaaa cttcttataaa ttctctctctc gacacttcgg taattcctct 60
ttcgagacta aagctctttt tgtatgcgtg tgtgtcaagc gtatgccccg ggattctcct 120
ccgcttcctt ttctcggtct tccttcttgc tttagggacc ggaagagtc ttgaaccaa 180
atagctcggc gggcacttcc ggggcccggc cccagagttc cgggaggggtg caggcaggag 240
agggaaagcg agcagcggcg gcagctggag gatgaagaag gagcatgtgc ttactgcca 300
gttctcccg tggtaccctt tcttccgagg cgttaccatc aagagtgtca ttctccact 360
tcctcagaat gtgaaggatt atttactcga tgatggaact ctggtgggtt caggaaggga 420
tgatccacca acacattctc agccagacag tgatgatgaa gcagaagaaa tacagtgggtc 480
tgatgatgag aacacagcca cgcttacggc accagaattt cctgagtttg ccactaaagt 540
ccaggaagct atcaattccc tcgggggcag tgtctttcct aagcttaatt ggagtgtccc 600
aagggatgcg tattggatag caatgaatag ttctctgaaa tgtaaaaccc tcagcgacat 660
ctttctgctt ttcaagagt ccgatttcat cactcgtgac ttactcagc cgtttattca 720
ttgtactgat gattctccag atccatgtat agaatatgag ctogttctcc gaaaatgggtg 780
tgaattgatt cctggggctg agtttctgat ttttgtcaag gaaaacaagc ttattgttgt 840
gttcgatata tacagagaca gtagggggaa ggtgtggctc attgacttta atccatttgg 900
tgaagtcaca gattcactgc tgttcacctg ggaagaactg atatctgaga acaacttaaa 960
cggcgatttt agtgaagttg acgctcaaga gcaggattcc ccagctttcc gttgcacaaa 1020
cagtgaagtg acagtccagc ccagccccta tttagttac cggctaccca aggactttgt 1080
agacctctct actggggagg acgctcaca gctaatagac ttcttaagc tgaagagaaa 1140
tcagcaggag gacgactgat gagcgtactg taactggaga agaggaggcc ccgcccacc 1200
gctccgggag ctgctcatca gccgcaactt cctgccgacc ctgatgcggg tgggcccagc 1260
agtgtggaca tcagccactt tttatattca tgtacattca cctggggaaa aaaacggagg 1320
gactttgcta cttgtaaaaa taacataata aatagatctt aaacatagga aaaccatact 1380
gttctgataa taaaatgctt tctatgaaat aaaaaaaaaa aa 1422

```

&lt;210&gt; 50

&lt;211&gt; 1822

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506235CB1

&lt;400&gt; 50

```

cgggtggcgga ggcgggtgagt gcgcggtctcc ggggctggcc gactccgcta gtggcccggc 60
cggcctgtgc tggggggctc cgggctctgg gctctgggtg cgcggaccgg gccaggctgc 120
ttgaagacct cgcgacctgt gtcagcagag ccgccctgca ccacctgtg catctcttc 180
tttaagtttg atcctcgccc tgtttccaaa aacgcgtaca ggctcatctt ggcagccaac 240
agggatgaat tctacagccg accctccaag ttagctgact tctgggggaa caacaacgag 300
atcctcagtg ggttgacat ggaggaaggc aäggaaaggag gcacatggct gggcatcagc 360
acacgtggca agctggcagc actcaccaac tacctgcagc cgcagctgga ctggcaggcc 420
cgagggcgag gcacctacgg gctgagcaac gcgctgctgg agactccctg gaggaagctg 480
tgctttggga agcagctctt cctggaggct gtggaacgga gccaggcgct gcccaaggat 540
gtgctcatcg ccagcctct ggatgtgctc aacaatgaag aggcgcagct gccagaccgg 600
gccatcgagg accagggtgg ggagtacgtg cagcccatgc tgagcaagta cgcggctgtg 660
tgctgtcgct gccctggcta cggcaccaga accaacta tcatcctggg agatgcggac 720
ggccacgtga ccttactga gcgtagcatg atggacaagg acctctcca ctgggagacc 780
agaacctatg agttcacact gcagagctaa cccacctct gggcctggcc agtgggctcc 840
tggggggccc tgcttgagg ggcactgtgg acaggaaacc ttcttttgc atactgcatt 900
gactgcccg tggcttggcc agcatcccc ggtcaggggc cctgtggttt gcgtgttacc 960
catctgtgtc cccatgccc gttcagggtc tgctttatg ccagtggga gcagcagagt 1020
ctgatactag gtctaggacc ggccgaggta taccatgaac atgtggatac acctgagccc 1080
actcttgcac atgtacacag gcactcacat ggcacacaca tacactcctg cgtgtgcaca 1140
agcacacaca tgcaagccat atacatggac accgacacag gcacatgtac gtgcacagg 1200
gtgctacaca tgtgcacaca tgcacagttg cacacacaga cacacacagg tgcacacaca 1260
cgatgccgaa caaggcagaa gggcgactct cacctctcat gtgctctgg ccagtaggct 1320
tttgttcttg tccaacgaca ggagtaggct tgtattttaa agcggccct cctctcctgt 1380
ggccacagaa cacaggcgtg cttggactct tgacaagcag acctgctcct gcagaggaga 1440

```



```

cagccacatt tgggaattggg caccgagaag acctgagaaa aacccactct ctcttttttt 1500
tttttttgag acggagtctt gctctgtcac ccaggctgga gtgcagtggc acgatctcgg 1560
ctcactgcaa cctccgcctc ccaagttcaa gcaattctcc tgccccagcc tctgattag 1620
ctgggattac aggcgtgtgc caccatgccc agctaatttt ggtgttttag tagagacggg 1680
gtttcaccat gttggtcagg ctggtcttga actcctgacg tggatgacca cctgcctcgg 1740
cctcccaaag tgctgggatt acaggcgtga gccaccacgc ctggccgaaa aacccactct 1800
cataacagac tttccagcac ag

```

1822

&lt;210&gt; 51

&lt;211&gt; 6104

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1302184CB1

&lt;400&gt; 51

```

gctgccggga catggttgca gatctgatct cttctgaaca cctcatcgtg tctccatccc 60
tgggaatctg accctagcaa ctggaccact ttgttcttgg aattttgggt gtcccttttt 120
ctcacttttc ctttttccct tttcccttc cccctcctga gaactccgga agactgtaga 180
gattgtcatg gagtcacagg aaaccttaag cagctcccgg caaagagggg gcgagtcaga 240
cttctgcccgt gtctcctcag ccaagcccc agctgtctct ggctgtgcag gagaaccttt 300
gctctccact ccaggacctg ggaaggggat cccggtgggc ggagagcgca tggagccaga 360
ggaggaggat gaactaggct cagggcgga tgtggattcc aactccaacg cggacagtga 420
gaaatgggtg gcaggagatg gtttgaaga gcaggaattt tctatcaagg aggcaaaact 480
cacagaggga agtctgaagc taaagattca gaccacaaag cgggctaaga aacccccaaa 540
gaatttgagg aactatatat gtccacctga gatcaagatc accatcaagc agtctgggga 600
ccagaagggtg tcccgtgctg gaaaaaatag caaagccacg aaggaggaag aaagaagcca 660
ctccaaaaag aagctcctca cagccagtga ccttgacgcc agtgacctca aaggatttca 720
gccacaggct tacgagaggc cccagaaaca ttcaactctc cattatgaca cgggcctccc 780
acaggacttc accggtgaca ccttaaaacc aaagcaccag caaaaaagca gcagccagaa 840
ccacatggac tgggtccacca actctgacag cggaccgctc actcagaatt gcttcacag 900
tccagagtct ggagagaaaa ctgcaagcac cagcaagatc cccgtctctg agcccgtggc 960
ttcctttgca aaggcccagg gtaagaaagg cagtgcaggg aacacgtgga gtcagtgtgc 1020
taacaataac aaagatctgc tcttgggagg tgtggctcca tccccagca gccacagctc 1080
accagcccca cccagcagct ctgctgagtg caacgggctt cagcccttgg tggatcaaga 1140
tggaggagggt acaaaggagc cccagaaacc acctacgggt ggagcaaga aaaagtccag 1200
taaaaaagat gtgataagtc agaccatacc aaaccagac ctggattggg tcaagaatgc 1260
ccagaaagca ttgacaata cagaagggaa aaggggaagg tattccgcag atagtgccca 1320
agaggcatca ccagccaggc agaacgtgag ttctgccagt aatcctgaaa atgactcaag 1380
tcatgtccgg attactatcc ccatcaaggc acctctctg gatccaacca accataagag 1440
gaaaaaaaga cagtccatta aagcgggtgt ggaaaagatc atgccagaga aagccttggc 1500
ttctggaatc accatgagca gtgaagtagt taacaggata ctttccaact ctgaggggaa 1560
taagaaggat ccccggtgtcc ctaagttgag taaaatgata gagaatgagt cccctcagct 1620
tggccttgaa actggtggaa atgctgagaa agttatccca ggagggtgtgt ctaagcccg 1680
gaagccaccc atggtcatga cacctccaac gtgcacagat cactctccat ccagaaagct 1740
gccagaaatc cagcatccaa aatttgctgc tggacttgca gcaaaccaaa 1800
acctagcacc atgcttcgag aggcagttat ggccacctct gataaactga tgtggagcc 1860
cccgctgca tatcccatca ccccatccag cctctcttac accaacacag acagtcttac 1920
tgtatcact ccagtcaaaa agaagcgggg acgaccaaaag aagcagcctt tgctcacagt 1980
cgagacgatt catgagggaa cttccaccag cccgctcagt cccatcagcc gagagtttcc 2040
tggcactaag aaaagaaagc gacgacgcaa tttagcgaag ttggcccagc tagtgccggg 2100
agaggacaaa cccatgagcg agatgaaatt tcacaagaaa gttggaaagc tcggcgtgtt 2160
ggataagaag accatcaaaa ctatcaataa gatgaagaca ctcaagagga aaaacatctt 2220
gaatcagatc ttgtcctgtt ccagcagcgt tgctctgaag gcaaaagctc cccagagac 2280
cagccctggg gcagcagcca ttgaaagcaa actgggcaag cagattaatg tcagcaagag 2340
gggaaccatc tacattggca agaagcgggg caggaagcca agagcagagc tgccaccccc 2400
atccgaagaa cccaaaacag ccatcaagca cccagccct gttctagcc agccggatgt 2460
tccagccgtg ccttccaact ttcagtcact tgtggcgtct tcaccagcag ctatgcacc 2520
actttcaaca cagttagggt ggtccaatgg caacctgagc cctgccagca ctgaaaccaa 2580
tttttcagag ttgaaaaacta tgccaaatct ccagcccatc agtgccttc caacccaaac 2640
ccaaaaggga atacacagtg gaacctggaa gctgtctcca cccagactga tggccaactc 2700

```



```

cccttcacac ctgtgcgaga ttggctccct aaaggaaatc acgctgtccc ctgtgagcga 2760
gtccacacagt gaggagacga tcccacgcga cagcggcatt gggacagaca acaacagcac 2820
ttctgaccaa gcggagaaga gctcagaatc ccgaaggagg tactcttttg atttctgttc 2880
cctggacaac ccggaggcca ttccgtccga caccagcaca aagaaccggc atggccaccg 2940
gcaaaagcat ctcatgttg acaactttct ggccacgaa agcctcaaga agccaaagca 3000
caagaggaaa cggaaaagcc tgcaaaaccg cgatgacctc cagtttctgg cagacctgga 3060
ggagctaata accaagttcc aagtgttcag aatctccacc cggagttaca ccttctacca 3120
cgagaatcca tatcccagca tttttcggat taattttgat cactattacc cgggtgccata 3180
tatccagtat gacccgttgc tctatcttcg taggacttca gacttgaagt caaagaagaa 3240
gcgtggtagg cctgcaaaaa ccaatgacac catgacaaaag gtgccttttt tacaagggtt 3300
cagctaccct attcccagtg gaagttacta tgaccctat ggaatgcctt acacatcaat 3360
gcctatgatg aaccttgggtt attacgggtc aatcccagct cctttgtacc tatcgacac 3420
gcttggagca gcttccccat tcatgaggcc aacagtgcc aacacctcag tccacacaaa 3480
ctcccacgta aagatgtccg gtgcagctaa gcataaagcc aagcatggag tacacctgca 3540
gggacctgtt agcatgggccc ttggtgacat gcagccttct ctgaaccctc ccaaggtagg 3600
cagtgccagt ctgtcccagtg gtcggctcca taaggagaaa cacaacacac agcataagca 3660
caaggaagac cggatcctag ggacctaga caacctgagt ggtttttttg caggcaaagc 3720
cacaggcttc tccagccaca tccgtgagcga gcggctgagt agcgcagaca aagagctccc 3780
gctggtgagt gagaagaaca agcataagga gaaacagaag caccagcaca gcgaagccgg 3840
ccacaagct tctaagaaca actttgaggt ggacaccctg tctacactgt cactttccga 3900
cgcccagcat tggacacagg ccaaggaana aggagacttg agcagtgagc ctgtggactc 3960
atgcacgaaa agatactctg gcagtggcgg gcatgggtgg agcacgagat cggagaacct 4020
ggacgtgttc agtgaaatga acccttcgaa tgacaagtgg gacagtgacg tgagtgggag 4080
taaaaggagg agctatgaag gctttggaac gtacagggaa aaggacatcc aagccttcaa 4140
gatgaaccgc aaggagagaa gttcttatga ctctccatg tctccaggga tgccaagtcc 4200
ccacttaaaa gtggaccaga cagcagtga tagtaagaac gaaggctcag tgcccaccat 4260
gatgaccagg aagaagccag ccgcagttga cagtgttaca attccaccag cccagtggtt 4320
atctctcctt gctgcatctg cagcaacgtc ggtgcagtc ggctcctccc tgaagaagag 4380
gttcaagcgg cgggagatcg aagccatcca gtgcgaagtg cgggaagatgt gcaactacac 4440
caagatcctg tccaccaaga agaacctgga ccacgtgaac aagatcctga aggccaaagc 4500
gctgcagaga caatcaaaaa caggcaacaa cttcgtgaag aagaggcgcg ggcgtcccag 4560
gaagcagccc acccagttcg atgaggactc cagagacca atgcccgtgc tggaaaaatg 4620
catcgacctg cccagcaaaa gaggccagaa gccacgctg agcccgtgg tgctggagcc 4680
cgccgccagc caagacacca tcatggccac catcgaggcg gtcattccaca tggcccggga 4740
agcgcgcgcc ctgccccgc caccgcgcgc gccctgcgc ccaccgcgc caccacct 4800
gccccgcga cccctctac ccaagacccc ccgaggcgga aagaggaaac acaaaccgca 4860
ggccccgcgt cagccccac agcagtgcgc cccgcagcag ccccttcccc aggaagagga 4920
ggtgaaagcc aaaaggcaga ggaagtccc agggagtga agcgagggtc ttccctaggg 4980
cgggtctggg cgtctgcacc tggggcctag ggaactgaca cgtgggaagc gcagtgaacc 5040
ggggcggggg cgggaatcccc cgctgcaggg acaccacgc ccttctctcc agaagccggg 5100
caggcagaat ccggccagac gacggggctg agccatcagg agctcttggg aaagcaaagc 5160
aggagacac cttcagaaga agcttgtctg agcttcaccg cagctccacc cagcgggcgc 5220
ttcagtacgg ctggatcctc cgaggcgag cggaaagccc ccaggaggag caggctgggtg 5280
gcactctcct gacctcacgc tgccaaggtg cgccctcgac tcccgaattc atttgetggt 5340
caggaaaaatc gaaagggaaa caccctcaat taggaaacat tccaagggga gaaaagctgc 5400
aaattgctca actggcattg ctgtaaccgg ttccatttaa ttgggttttt tttcttttga 5460
ttttcaagct ctgctaagct ttgggggtacc aacgtcatct tcaggtgacc tgaatctttc 5520
ccttaaccgt acagtttctc gatggaattg tgtgatcaga aggtggaatt ctagtgatag 5580
gcgacctcag acccgcatte atgttctgtg tgccctctct attgcacata cactgatttt 5640
tagcattgtc tattcctatt tttcctttgc ccattgtact tccatataat ttttcattaa 5700
cttacttgct gccttttttt tttcttggta cacatttaaa taaagtaatc cttaacctgt 5760
gctgtaaagt tcacccttgg catgctgttc caagaacctg ggtttgaatc ccaatcggtg 5820
tgaaacatac tcagtattga taaaaccttt ttaataagtg atgcagagca gccaaaggata 5880
tggtgaccca gatgtcaacc aggtattttt tataactaaa acatgtcagc agagcatagg 5940
cagaataaaa tggtttaaat accccacagc aaatagagta actgacaaac cacaaaaaac 6000
tgaaacccca gaccaccag aaagacaagt gtctagcaat gccttggtac ctgatctttt 6060
tcattcaaat aattgtcata ggtgttcaa aaaaaaaaaa aaaa 6104

```

&lt;210&gt; 52

&lt;211&gt; 1137

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506232CB1

&lt;400&gt; 52

```

ggagtgggag cgagcggcgg ggccagctgc gttctgagcc tgggcgccagc tgccatctgc 60
tctgggaagc accaggggtgt ccccgccgcc ctgagctcga agtcagccac catggaggcg 120
caggcacaag gtttgttgga gactgaaccg ttgcaaggaa cagacgaaga tgcagtagcc 180
agtgtctgact tctctagcat gctctctgag gaggaaaagg aagagttaaa agcagagtta 240
gttcagctag aagacgaaat tacaacacta cgacaagttt tgtcagcgaa agaaaggcat 300
ctagttgaga taaaacaaaa actcggcatg aacctgatga atgaattaaa acagaacttc 360
agcaaaagct ggcatgacat gcagactacc actgcctaca agaaaaacaca tgaaaccctg 420
agtcacgcag ggcaaaaggc aactgcagct ttcagcaacg ttggaacggc catcagcaag 480
aagttcggag acatgaggaa ttctcctact ttcaaatcat ttgaggagag ggttgagaca 540
actgtcacaa gcctcaagac gaaagtaggc ggtacgaacc ctaatggagg cagttttgag 600
gaggtcctca gctccacggc ccatgccagt gccagagct tggcaggagg ctcccgccgg 660
accaaggagg aggagctgca gtgctaagtc cagccagcgt gcagctgcac ccagaaaccg 720
gccactacc agcccatctc tgctctgtct tatccagata agaagaccaa aatcccgtcg 780
ggaaaaacc aggcccttgac attgttatcc aaatgggccc tccagaaagt ttaatgattt 840
ccatcttgta tttgtgttga tgatggagca ctgggacatc acattttcag tttcacagat 900
gactgtcaca tttttaaattg ttccccttga gcaggtccac aacgggtcta attcccgtcg 960
tggaattcg tggtaatattt ttcaaacatg tagcaatggt ttgcttgatt tttgcctggc 1020
tcccttatga agggcaggcc ctgaagggtg attgaaaggc ctttccgttc gcgattaaaa 1080
gatggggcct tccgacggtt tccgcaatag atggttcctt aaattcatcc accctgt 1137

```

&lt;210&gt; 53

&lt;211&gt; 1011

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2585358CB1

&lt;400&gt; 53

```

gtagggcgga gctcgcgggg tgccgctctg ggataggcga ccacggtgtc ttcaaaagcc 60
ccgtcagggg tggcttccctg gggccggacc gactgtgggt cagtttgac cagcgtctctg 120
gaatcgaggt acgcgcgaaa gggcagagtt tctggaggaa accgcagcct ctcaaccgct 180
gaccgggtct cagaaggccc ccggcagggc cgcttggcgg gaactgacca cgcgccagtc 240
aggctctcca gggacctgcg caggcgcggtg tgggaggagt cgtgcgcagg gggcggggct 300
tcgggaagga gccacagaga gggcggggcg taggacctgc gcttcggggg tggagtccga 360
gcggcgccgg gcgggtcatg cgggacgcgg atgcagacgc aggcggaggc gctgacggcg 420
gggatggccg ggggtggccac agctgccgcg ggggcgtgga cacagccgca gctccggccg 480
gtggagctcc ccagcgcac gcgccaggtc cgggcagaga cgcgcgctct gcggccaggg 540
ggtcacgaat gcggccgcac atattcacc tccagcgtgcc ttccccgacc cccttgagg 600
cggaatcgc ccatgggtcc ctggcaccag atgccagacc ccaccaaagg gtggttggga 660
aggatctcac agtgagtggc aggatcctgg tcgtccgctg gaaagctgaa gactgtcgcc 720
tgctccgaat ttccgtcatc aactttcttg accagcttc cctggtggtg cggaccatgc 780
agcgttttg gcccccggt tcccgctaag cctggcctgg gcaaatggag cgaggtccca 840
ctttgcgtct ccttgtaggc agtgctcca tccctcccta gggcaggaat tcccacagtt 900
gctactttcc tgggagggcc tcatgtttta tctggttctt aaatgtttgt tactacagaa 960
aataaaactg cgctactaaa aaaaaaaaaa aaacctcgtg ccgaattctg t 1011

```

&lt;210&gt; 54

&lt;211&gt; 4638

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3961495CB1

&lt;400&gt; 54

gaccctttga aactgtcgag tgacctggga tccgtttaaa cactgacctg cttttagtaa 60  
 gaaatccaac cctatcaatg caatgcagac aaatggactg ccaaactggg gcatggctgt 120  
 tggatatggg gccatcatcc cgccccacct gcaaggcctt cctgggtgcta atggatcatc 180  
 agtttctcaa gtcagtggg gtcagtgtga aggaataagc aattctgtgt ggggactgtc 240  
 cccaggtaac cctgccacag gaaatagcaa ttctgggttc agtcagggga atggagacac 300  
 tgtgaactca gcattaagt ctaaacaaaa tggatccagc agtgctgtgc aaaaggaagg 360  
 aagtggagga aatgcttggg attcaggacc tctgtctggt cctggaatac tcgcttggg 420  
 aaggggcagt ggcaacaatg gcgttggtaa tatccattca ggagcttggg gccacccacg 480  
 ccgaagcacc tctaacggtg tgaatgggga atggggaaag ccccaaaacc agcattccaa 540  
 cagtgcacac gatcgggaaag gatcaacagc gtgggagagt cctagtgtca ccagccagaa 600  
 cctaccgta cagcctggtg gtgaacacat gaactcctgg gccaaagcgg catcttcttg 660  
 aactacagca agtgaaggaa gtagtgatgg ttctggcaac cacaatgaag gaagcactgg 720  
 gaggggaagga acggggagaag gccgaaggcg agataaaggg attatagacc aagggcacat 780  
 ccagtgtcca aggaatgatc ttgacccaag agttctgtct aatactggtt ggggacagac 840  
 tctgttaaag caaaacactg cctgggaatt tgaagaatcc cctaggctctg aaaggaaaaa 900  
 tgacaatggg acagaggcct ggggttgtgc agtactcag gcttcaaaact caggggggaa 960  
 gaacgatggg tccatcatga acagtacaaa taccctctca gtatctgggt ggtcaacgcg 1020  
 gccacctgcc gctgtgccag caaacacagg ttggggagac agcaacaaca aagcgccaag 1080  
 tggcccgggg gtttgggggg actcgataag ctctactgct gttagtactg ctgctgctgc 1140  
 caagatggc catgcttggg gtggggccgc aaatcaggag gacagatcaa atccagcctg 1200  
 tgagcctcca aagcccaaat cccaacactg gggagatgga ctaaggactc tgggggatgg 1260  
 gagtgcagga gggggagatt gggcagattc atcgctgtgc cagggtctg gttggaatga 1320  
 gaaaaaaaat ggtcttggat gggatgtctg cagtaatagg tcagggtctg gttggaatga 1380  
 caccacgaga tctgggaaca gtggctgggg caacagcaca aatacaaagg ccaatccagg 1440  
 tacaactgg ggggagactt taaaacctgg cccccaacag aactgggcta gcaaaccaca 1500  
 agacaacaat gtgagtaact ggggaggagc tgcttctgtg aaacagacag gaacagggtg 1560  
 gatcgggggg ccggtaccgg tcaaacagaa ggacagcagt gaagcaactg gctgggaaga 1620  
 accctctcca ccgtccattc gccgcaaaat ggaaattgat gatgtaacct cagcttgggg 1680  
 ggacccaagc aactataaca ataaaactgt aaacatgtgg gatagaaaca acccggtcat 1740  
 ccagagcagt accacgacca ataccaccac caccaccacc actaccacga gcaacaccac 1800  
 acacaggggt gagacgcgc ccccgacca ggccggtact cagctgaatc gatcaccgtt 1860  
 gcttgggtcca gtttcatcag gctggggaga aatgcctaag gttactcaa agactgaaaa 1920  
 ctcttgggga gaaccatcct ccccttctac cctggtgat aatggcacag cagcatgggg 1980  
 gaagccaccc agcagtggca gcgggtgggg agatcacctt gcagagccgc cggtggcatt 2040  
 tgggaagagc ggcgcacctg ttgctgcctc agcctgtgac aaaccagctt caaaatctat 2100  
 gcaagaaggc tggggcagtg gtggggatga aatgaacctc agtaccagcc agtgggagga 2160  
 tgaagaaggg gacgtgtgga ataagtctgc ttccaagaa agcacctcct cctgcagctc 2220  
 ctgggggaac gcccccaaaa aaggacttca aaagggcata gacatgggct tcccgagaga 2280  
 tgaggcctgg atcatgagcc ggctgatcaa acaactcaca tatgaatctt gatcaggcca tgagcgctct 2340  
 gccagctgag gaggccttga agagtaacaa tgggctggga gtagccgacc ataattggaat 2400  
 gctggaaaag aaggtggacg ggcgcccgc aatctccaaa agtcttccg tggaccgccc 2460  
 ggcgccaag cccctcggtc gacaaggatg ggcgcccgc ggaagagccc acgcttccac cgttcttggc 2520  
 cacttttctt ctgaagctcc ccctttcaca cagtgcactc cccagtcagg ccctgggtgg 2580  
 gattgcctcc gggctgggca tgcaaaaact gaattcttct agacagatac cgagtggcaa 2640  
 tctgggtatg tttggcaata gtggagcagc acaagccagg accatgcagc agccgccaca 2700  
 gccaccagt cagcctctta cagcctttt gcaagcttca gctgtcaag tgcctcagtt 2760  
 tctatccctt caggttcaag caatattacc tcaacatatg acgatgttga accagctcta 2820  
 ccctgcacta ttaacctcgc aacgtttaca aatccagcag cagatgttac agggccagcg 2880  
 tcagctgcag ctggcatacc gacacagga gcagcaagtt gcgcccacaa tcaactatct 2940  
 taatgtgtcc ggatccatga gacacagga gctggcccag gccctgctcg tgaagcagcc 3000  
 gcagcagcag atccagcagc accagcgcca gctggcccag tctgcacccc tctgcaggca aatcgccat 3060  
 accaccgcca ccgcccgcg cgacactgtc cctgtcactc gagctgcaga ccaaagagca 3120  
 ggacagcttc cctcgcacc cacagactcc cggcctacct gacttgaaac caaacatgaa 3180  
 gcagtcttca cccaacacct ttgctcctta cctctcgtc ggactgaacc agtcccagtc 3240  
 tgtaacagc atggacatga ccggtggctt gtcggtgaag gacccatccc cgtcttcccc 3300  
 acgctcccc cagtggagc accccaactc catggataac ttgcccagtg ccgcttcccc 3360  
 cctggagcag aacctagca agcatgggtg tatccctgga ggtctaagca ttgggctccc 3420  
 aggtaagtcc tccattgatg actcctatgg ccggtacgat ttaattccaga acagtgaagtc 3480  
 accagccagt cctccgtag ctgttcccc tagccggtca cgtgccaaat ctgacagtga 3540  
 taaaatctca aatggcccta gcatcaactg gccccagaa ttccatccgg gacttccatg 3600  
 gaaaggactg cagaatattg accctgagaa tgaccctgac gtcactcctg gcactgtccc 3660  
 cactgggcct accatcaaca ccaccatcca ggatgtcaac cgctacctcc tcaagagtgg 3720

```

aggtaaactg tcagacatca aatcgacgtg gtcctctggc cctacctccc acacgcaagc 3840
ctctctgtct catgaactat ggaaggtgcc cagaaacagt actgcaccca cgaggccacc 3900
tccagggtta accaatccca agccctcctc cacctggggg gccagccccc tcggctggag 3960
cagctcctac tccctgggtt ctgcctggag caccgacacc tcaggaagaa ccagcagctg 4020
gctcgttctt cgaaacctca ctocccagat tgatgggttct aactgcgga cattgtgttt 4080
gcaacatggg cctcttatca cattccacct gaatctgact caaggcaatg ctgtgggtccg 4140
gtacagctcc aaggaggagg ctgccaaagg ccagaagtct ctgcacatgt gcgtcctggg 4200
aaacactacc atcctggccg agttcgtctg tgaagaagaa gtgaatcgct tcttagccca 4260
aggccaggcg ctgccaccca cttccagctg gcagtcacgc agcgcgtcca gccagcccg 4320
gctcagcgca gcgggcagct cccatggcct ggtacgcagc gacgctggcc actggaacgc 4380
cccgtgcctg ggtggcaagg ggagcagtga gctgctgtgg ggccgggtgc cccagtaact 4440
cagcagcctg tggggcccg ccagcgccga cgacagcagg gtgataggca gcccacgcgc 4500
gctaaccacc ctgctgcctg gggacctgct cagcggggag tccctgtagg ctctgccatc 4560
atcagcacca ggagagccga cccctcccgg gacctctccc ggctggggcg cccacagac 4620
ccgctggaac ccagcagc                                     4638

<210> 55
<211> 2233
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500801CB1

<400> 55
ctgccgggtg tgccgggtgt ccagcgaacc cttttcccaa accttcgggg agaagggagg 60
tgggaggagg caaagaaact acaggcaggg agctggaagg gggggtgggg ggggcaggag 120
acaagaaatc aagacaccag gcagcaggac acacacacac tcacatacac tcacacacat 180
agagaccaac agatagacag ctacctaaag cctgaaagac tgacagcaac acagaaaaaa 240
agaaacaggc agaaagagag acaaagacag aaatagaaac agactaacac acagagtcaa 300
aaatacagag acagaaagac agggagaaag agaaacagaa aattagacac caaagacata 360
cgaacaggga ggaaggccga ctgaaagaaa gacggagaag aggagagaga agccagggcc 420
gagcgtgcca gcaggcggat ggaggcgccg ctggtggagg aggagacgta gtggcctggg 480
ctgagctggg tgggcccggg gaagcgggtg cctcagagtg ggggtggggg catgggaggg 540
gcaggcattc tgctgctgct gctggctggg gcgggggtgg tgggtggcctg gagaccccca 600
aagggaagt gtcccctgcg ctgctcctgc tctaaagaca gcgccctgtg tgagggtctc 660
ccggacctgc ccgtcagctt ctctccgacc ctgtgtcac tctactcgt caggacggga 720
gtcaccacgc tgaaggccgg cagcttctct agaattccgt ctctgcacct gctcctcttc 780
acctccaact ccttctccgt gattgaggac gatgcatttg cgggcctgtc ccacctgcag 840
tacctcttca tcgaggacaa tgagattggc tccatctcta agaatgcctt cagaggactt 900
cgctcgctta cacacctaag cctggccaat aacctcttgg agacctccc cagattcctg 960
ttccgaggcc tggacacct tactcacgtg gacctccgag ggaacctgtt ccagtgtgac 1020
tgccgcgtcc tctggctcct gcagtggatg cccaccgtga atgccagcgt ggggaccggc 1080
gcctgtgcgg gcccgcctc cctgagccac atgcagctcc accacctcga cccaagact 1140
ttcaagtga gagccataga gctgtcctgg ttccagacgg tgggggagtc ggcactgagc 1200
gtagagccct tctcctacca aggggagcct cacattgtgc tggcacagcc cttcgccggc 1260
cgctgcctga ttctctctg ggactacagc ctgcagcgt tccggcccga ggaagagctg 1320
cccgcggcct ccgtggtgtc ctgcaagcca ctggtgctgg gcccagcct ctctgtgctg 1380
gctgcccgcc tgtggggggg ctcacagctg tgggcccggc ccagtcccgg cctgcgcctg 1440
gcccacacgc agacctggc ccgcggcggt ctgctgcggc ccaatgacgc cgagctcctg 1500
tggttggaag ggcaacctg cttcgtgggt gccgatgcct ccaaggcggt cagcaccacg 1560
ctgctgtgcc gcgacggggt cggcttttac ccgcaccaga gcctgcacgc ctggcaccgg 1620
gacacggacg ctgaggccct ggagctggac ggccggcccc acctgctgct ggccctcggt 1680
tcccagcggt ccgtgctctt ccactggacc ggtggcgct tcgagagacg caggacatc 1740
cccaggcggt aggatgtcta tgccacacgc cacttccagg ctggtgggga cgtgttctg 1800
tgccctcacac gctacattgg ggactccatg gtcattgcgt gggacggctc catgtttcgt 1860
ctgctgcagc aacttccctc gcgcgggtgc caccgtccca gcctactgga cctcccagac 1920
tgcccttcag gtctaaagca acatctggac atctcggctg gggcagattc cccatcaaag 1980
gcaacctggg ggggtgcaggc tgggtgcactg catggaggt aggtggaggt cagctgcaga 2040
gctccatgag gtcccggagc cccctctca acccgctcct agccccgtg gtccctctga 2100
aagcaccgag gggttgatgg cattccttcc ccatctctaa ctccacgcag ccaatcgctt 2160
tcctggagga gcgccaacct cgaagacaac tcagactgg gcgggttccc ctctgctgct 2220

```



tgtgcgcatg cgt

2233

&lt;210&gt; 56

&lt;211&gt; 4884

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506414CB1

&lt;400&gt; 56

```
atgctgaagt gcatcccgct ctggcggtgt aaccggcacg tggagtcggt ggataagcgg 60
cactgctcgc tgcagggtgt gccagaggag atttaccgct acagtcgtag cctcagaggag 120
ctgctcctcg acgccaacca gctgcgcgag ctgcccacgc cttttttccg gctgctgaac 180
ttgagaaagt tgggcctcag tgacaacgag atccagcggt tgccctccga ggtggccaac 240
ttcatgcagc tgggtggagct ggacgtgtcc cggaacgata tccttgagat cccggagagc 300
atcaagttct gcaaggctct ggagatcgcg gacttcagcg ggaacccccct ctccaggctc 360
cctgatggct tcactcagct ggcgagcctg gctcacctgg cctgaatga tgtgtctctg 420
caggcactgc ccggggacgt gggcaacctc gccaacctgg tgacctgga gctccgggag 480
aacctgctca agtcctcgcc agcgctccctg tcatttctgg tcaagctgga acagctggat 540
ctggggaggca acgatctgga agtgctgcca gacactctgg gggctctgcc caatcttcgg 600
gagctgtggc ttgaccggaa ccagctgtca gcaactgccc cggagctcgg gaacctgcgg 660
cgctgtgtgt gcctggacgt gtcggaaaac cggttgagg agctgcctgc tgagctcggc 720
gggtgtgtgc tgtcactga cctgctgctg tcccagaacc tgctgcggag gctgcccagc 780
ggcatcggtc agctgaagca gctatccatc cttaaagtag accagaatcg gctgtgcgag 840
gtgactgagg ccacgggga ctgtgagaac ctgactaagc tgatcctcac ggagaacctg 900
ctgatggccc tgccccgctc cctgggaaag ctgactaagc tgaccaacct caacgtggac 960
cggaaccacc tcgaggcgct gccgcccag atcgggggct gtgtggcact cagcgtcctc 1020
tccttgaggg acaaccgcct ggccgtcctg ccaccagagc tggcccacac gacagagctg 1080
cacgtgctgg acgtggcggg gaaccgcctg cagagtctgc cgttcgcgct caccacctc 1140
aatctcaagg cctgtgtggt ggcagagaac caggcgagc ccatgctccg gttccagacg 1200
gaggatgatg cccggaccgg cgagaaggtg ctacactgct acttgctgcc ccagcagccc 1260
ccactcagcc tcgagggatgc tgggcagcag gggagcctct cggagacctg gagcgatgcc 1320
ccgcccagcc cgtcagcgt ccatcagttc ctggaggccc ccataggtga tgaggacgct 1380
gaggaagctg cagctgagaa gcggggccta cagcgccggg ccacacctca cccagcgag 1440
ctcaaggtga tgaagaggag catcgagggg cggcggagcg aggctgccc ttgccagcca 1500
gactctgggt cgcccttgcc tgcaaggag gagaagcggc tgagtccga gtctggcctg 1560
agtgaagact ctgcgccatc tgccagcaca gtctctgagg ctgagcccga gggcccgctg 1620
gctgaggcac aggggtgggag ccagcaggaa gccacgactg ctggcgggga ggaagacgcc 1680
gaagaggact accaggagcc cacggtgcat ttcgcagagg acgcaactgct gcccggggat 1740
gacagggaga tcgaggaggg gcagccttag gccccctgga ccttgccagg cgggaggcag 1800
cggctcatcc gcaaggacac acctcactac aaaaagcact tcaagatctc caagctgccc 1860
cagcccagag ccgttgtggc tctgctgcag ggcattgcagc ctgatgggga gggccctgtg 1920
gctcccgggg gctggcaciaa tggcccccac gcaacctggg ctccctgggc ccagaaggag 1980
gaggaggagg aggaagagg tagtccctcag gaggaggagg aagaggagga ggaggaaac 2040
agggctgaag aggaagagg cagcactgag gaggaggaca aggagggggc cgtggtttct 2100
gcgccctctg tcaagggagt gtcgtttgac caggccaata acctgctgat agagcctgct 2160
cgcattgagg aggaagagct gacctcacc atcctgcggc agactggggg cctgggcac 2220
agcattgcgg gcggcaaggg ctccacaccc tataaggggg acgacgagg catattcatc 2280
tctcgggtgt ccgaggaagg ccctgcggcc cgggctggag tccgtgtggg tgacaagctc 2340
ctggagggtg atggtgtggc tctgcagggc gccgagcacc acgaggccgt ggaggcgctc 2400
cggggggccg gcaactgcgt gcagatgcga gtgtggcggg agcgcatggt ggagcctgag 2460
aacgcgggtc ccatcacgcc gctgcggccc gaggatgatt acagccccc agagcggcgg 2520
ggaggggggc tgcgctgccc cctgctcccc cctgagagcc ccggggccct ccgtcagcgc 2580
cacgtggcct gcctggcacg cagcgagagg gggctgggct tcagcattgc tgggtgggaa 2640
ggctccacac cctacagggc tggtgatgcg ggcattctcg tctcccgcac tgccgagggc 2700
ggtgctgctc accgcgcggg cacactgcag gttggcgacc gcgtctctc tgagatccgt 2760
ctgccaagag ctgggggccc tctggggctt agtattgtcg gaggctccga ccattccagc 2820
caccgctttg gtgtccagga gcctggtgtg ttcatctcca aggtgctccc gcggggcctg 2880
gccgctcgca gcggcctgcg ggttggggac cgcattcctg cagtgaacgg gcaagacgtg 2940
cgggatgcca cgcaccaaga agcagtcagt gccttgcctc ggccctgcct ggagctgtcg 3000
ctgctggtgc ggagggaccc ggcacccccg ggcctacggg aactgtgcat ccagaaggca 3060
```

```

cctgggggaga ggctgggcat cagcatccgc gggggtgccca gggggccacgc tggcaacccc 3120
cgcgaccccca cagacgaggg catcttcatc tccaaggtga gccccacggg ggacagcggg 3180
cgcgacgggtc ggctgcgtgt ggggtttgcgg ctgttgaggag tgaaccagca gagcctgctg 3240
ggcctgacgc acggcgaggg ggtgcagctg ctccgcagtg tgggagacac cctcaccgtg 3300
ctgggtctgtg acggcttcga gggcagcacc gacgcagccc tggaggtgtc cccaggtgtc 3360
attgccaacc cctttgcggc aggcctcggc caccggaaca gcctggagag catctcttcc 3420
atcgaccggg agctgagccc tgaggggccca ggcaaggaga aggagctgcc tggacagacc 3480
ctgcactggg ggcccagggc cacagaagcc gcaggtcggg gtctgcagcc cctgaagctg 3540
gactaccggc ccctggccgc cgtgcccagc gctggcagcg tgcagagggt accgtctgga 3600
gcagctggag ggaagatggc tgaatctccc tgctccccta gtggccagca gccgcccctc 3660
ccgctctctc cggatgagct gcccggcaat cctgcccagc cctacagggc cttcgcgggc 3720
gtgcccactt ctaccccgcc tgaggatgcc cctgcccagc cccccacgcc tgggacctga 3780
gcctccccgg agcagctgtc cttccgggag cggcagaagt actttgagct ggaggtgcgc 3840
gtgcccaggc ccgagggccc ccctaagcgc gtgtccctgg tgggtgctga cgacctgcgg 3900
aagatgcagg aggaggaagc cagaaaacta cagcagaaga gagcgagat gctgcgggag 3960
ggcgagagg ctggggccga agcgaggctc gccctggacg gggagacgct gggcgaggag 4020
gaacaggagg atgagcagcc accctggggc agcccagacc ccacctcaag gcagagcccg 4080
gcgctcccccc cgcccctggg aggtggcgcc ccggtgcgga cggccaaagc tgaacggcgc 4140
caccaggagc ggctgcgcgt gcagagtccg gagccaccgg caccgagcg tgccctgtcc 4200
cctgcccagc tccgggccct ggaggccgag aagcgtgcgc tgtggagggc agccaggatg 4260
aagtcatagg aacaggacgc tctccgagca cagatggctc tcagcaggtc ccaggaaggc 4320
cggggcacgc gggggccccc ggagcgactg gccagggccc cttcccctgc gccaccccc 4380
tcgcccaccc ctgtggaaga cctcgggccc cagaccagca cctccccggg acgctgtca 4440
ccggactttg ctgaggagtt gaggtccctg gaacctctc ccagccctgg cccgcaggag 4500
gaggatggag aagtggctct ggtgcttctg ggcaggccct caccggcgcc tgtggggcct 4560
gaagatgtgg cactgtgcag cagccgcgcg ccgtaaggc ctggggcgcc tggcctgggc 4620
cctgtgccct cctagaggag caggcacctc ccccagactt ggggtggggg ccctgccagc 4680
tccagacca cccttgccc aagtctttta acctgggtgt tagcatttta aagagacccc 4740
acaggagtgc tggcctgtga ctaactaact gccccacccc agccgagacc tcggcgagac 4800
tgtaactagt gatgttgta caaccaaa aagcctatctt tggtttaagg agaataaagt 4860
tgactacatt ttaaaaaaa aagg

```

&lt;210&gt; 57

&lt;211&gt; 5079

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7506415CB1

&lt;400&gt; 57

```

atgctgaagt gcatcccgtc ctggcggtgt aaccggcacg tggagtcggt ggataagcgg 60
cactgctcgc tgcaggttgt gccagaggag atttaccgct acagtcgtag cctcagggag 120
ctgtccctcg acgccaacca gctgcgcgag ctgcccgaag cttttttccg gctgctgaac 180
ttgagaaagt tgggcctcag tgacaacgag atccagcggc tgccctccga ggtggccaac 240
ttcatgcagc tgggtgagct ggacgtgtcc cggaacgata tccctgagat ccgggagagc 300
atcaagttct gcaaggctct ggagatcgcg gacttcagcg ggaacccctc ctccaggtc 360
cctgatggct tcaactcagc gcgcagcctg gctcacctgg cctgaatga tgtgtctctg 420
caggcactgc cgggggacgt gggcaacctc gccaacctgg tgacctgga gctccgggag 480
aacctgctca agtccctgcc agcgtccctg tcatctctgg tcaagctgga acagctggat 540
ctgggaggca acgatctgga agtgctgcca gacactctgg gggctctgcc caatcttcgg 600
gagctgtggc ttgaccggaa ccagctgtca gcaactgccc cggagctcgg gaacctgcgg 660
cgcttggtgt gcttgacgt gtcggaaaac cggctggagg agctgcctgc tgagctcggc 720
gggctggtgc tgcctactga cctgctgtc tcccagaacc tgctgcggag gctgcccagc 780
ggcatcggtc agctgaagca gctatccatc ctaaggtag accagaatcg gctgtgcgag 840
gtgactgagg ccacgggga ctgtgagaac ctctctgagc tgatccctac ggagaacctg 900
ctgatggccc tgcccgcctc cctgggaaag ctgactaagg tgaccaacct caactgggac 960
cgggaaccacc tcgaggcgct gccgcccagc atcgggggct gtgtggcact cagcgtctc 1020
tccttgaggg acaaccgcct ggccgtctct ccaccagagc tggcccacac gacagagctg 1080
cagctgctgg acgtggcggg gaaccgcctg cagagtctgc cgttcgcgct caccacctc 1140
aatctcaagg cctgtggct ggcagagaac caggcgagc ccatgctccg gttccagacg 1200
gaggatgatg cccggaccgg cgagaagggt ctcacctgct acttgcctgc ccagcagccc 1260

```

ccactcagcc tcgaggatgc tgggcagcag gggagcctct cggagacctg gagcgatgcc 1320  
 ccgcccagagcc gcgtcagcgt catccagttc ctggaggccc ccataggtga tgaggacgct 1380  
 gaggaagctg cagctgagaa gcggggccta cagcgccggg ccacacctca ccccgagcag 1440  
 ctcaagggtga tgaagaggag catcgagggg cggcggagcg aggcctgccc ttgccagcca 1500  
 gactctgggt cgcccttgcc tgcagaggag gagaagcggc tgagtgcga gtctggcctg 1560  
 agtgaagact ctcgcccatc tgccagcaca gtctctgagg ctgagcccca gggcccctcg 1620  
 gctgaggcac aggggtgggag ccagcaggaa gccacgactg ctggcgggga ggaagacgcc 1680  
 gaagaggact accaggagcc cacggtgcat ttgcagagg acgcactgct gcccggggat 1740  
 gacagggaga tcgaggagg gcagcctgag gcccctgga ccctgccagg cgggaggcag 1800  
 cggctcatcc gcaaggacac acctcactac aaaaagcact tcaagatctc caagctgcc 1860  
 cagcccagag ccgttgtggc tctgctgcag ggcatgcagc ctgatgggga gggccctgtg 1920  
 gctcccgggg gctggcaca tagtccctcag gacccctggg ctccctcggg ccagaaggag 1980  
 gagggaggag aggaagagg cagcactgag gaggaggagg aagaggagga ggaggaaaac 2040  
 agggctgaag aggaagagg cctcaccatc ctgcccaga ctgggggccc ggtggtttct 2100  
 gcgcccctct tcaagctgac cacacctat aagggggagc acgagggcac attcatctct 2220  
 attgcccggc aggaaggccc tgccggcccg gctggagtc gtgtgggtga caagctcctg 2280  
 cgggtgtccg gtgtggctct gcagggcgcc gagcaccacg aggcctgga ggcgctccg 2340  
 gaggtgaatg ctgccgtgca gatgcgagtg tggcgggagc gcatggtgga gcctgagaac 2400  
 ggggcccggca ctgccgtgca gcggcccag gatgattaca gcccgcgaga gcggcgggga 2460  
 gcggtacca tcacgccgct gctcccgcct gagagccccg ggcccctccg tcagcgccac 2520  
 ggggggctgc gctgcccct egagaggggg ctgggcttca gcattgctgg tgggaaaggc 2580  
 gtggcctgcc tggcacgcag tgatgcgggc atcttctct cccgcatgct cgagggcgt 2640  
 tccacacct acagggtg actgcaggtt ggcgaccgct tctctctat taatggagt 2700  
 gctgtcacc gcgcgggac agccaaggc gtctccctgc tgaccgctgc cccccacc 2760  
 gacgtgact agccaaggc tgaccagcc ggcctcttc ctccagccc tetgccacat 2820  
 atcgccctgc tggggagcg ggaggctgg cccctcttc ccccggggtg 2880  
 tctcaccac ccaccgctgc tgttgccacc accagcataa ccaactgccac ccccggggtg 2940  
 cctgggttgc cgagcctggc cccagcctg ctggctgccc cgttggaagg gccatacca 3000  
 gtggaggaga tccgtctgcc aagagctggg ggcctcttg ggcttagtat tgtcggaggc 3060  
 tccgaccatt ccagccacc gtttggtgct caggagcctg gtgtgttcat ctccaagggt 3120  
 ctcccgggg gcctggccgc tcgcagcgcc ctgcccgttg gggaccgcat cctggcagtg 3180  
 aacgggcaag acgtgcggga ggtgcggagg gacccggcac ccccgggcct acgggaactg 3240  
 tgcttgagc aggcacctg ggagaggctg ggcacagca tcccgggggg tggcagggg 3300  
 tgcattccga acccccgca cccacagac gagggcattc tcatctccaa ggtgagcccc 3360  
 cagctggga ccgggcccga cggtcggctg cgtgtgggtt tgcggtgtt ggaagtgaac 3420  
 acgggggag cagcagagcc tgctgggccc gacgcagggc gaggcggtgc agctgtccg cagtgtggg 3480  
 gacacctca ccgtgctggt ctgtgacggc ttgcaggcca gcaccgacgc agccctggag 3540  
 gtgtccccag gtgtcattgc caacctctt gcggcaggca tcggccaccg gaacagcctg 3600  
 gagagcatct cttccatga ccgggagctg agccctgagg gccaggcaa ggagaaggag 3660  
 ctgectggac agacctgca ctgggggccc gaggccacag aagccgcagg tcgggggtctg 3720  
 cagccctga agctggacta ccgcccctg gcccgctgc ccagcgtgg cagcgtgcag 3780  
 agggtaaccg ctggagcagc tggagggaag atggctgaat cctcctgct cctagtggc 3840  
 cagcagccgc cctcccggc tctctccgat gagctgccc ccaatgtgaa gcaggcctac 3900  
 agggccttcg cggccgtgccc cacttctcac ccgctgagg atgcccctgc ccagccccc 3960  
 acgcctggg ctgcagccct cccggagcag ctgtccttc gggagcgga gaagtacttt 4020  
 gagctggagg tgcgcgtgcc ccaggccgag gcccctcta agcgcgtgc cctggtgggt 4080  
 gctgacgacc tgcggaagat gcaggaggag gaagccagaa aactacagca gaagagagcg 4140  
 cagatgctgc gggaggcggc agaggttgg gcccgaagca ggctcgccct ggaaggagg 4200  
 acgctgggag agggaggaaca gggagctgag cagccacct gggccagccc gagccccacc 4260  
 tcaaggcaga gcccggcgtc ccccccggc ctgggaggtg gcgcccgggt gcggaaggcc 4320  
 aaagctgaac ggcgccacca ggcgtgcaga gctcgagag ccgagaagcg tgcgctgtgg 4440  
 gagcgtgccc tgtcccctgc cgagctccgg gcccctgagg gacacagat ggtcctcagc 4500  
 agggcagcca ggatgaagtc attggaacag gacgtctct cactgggagc ggtccccttc 4560  
 aggtcccagg aaggccgggg caccgctgtg gaagacctcg gcccagac cagcactcc 4620  
 cctgcgccc cccgctgccc ctttctgag gagttgaggt ccctggaacc atctccagc 4680  
 cgggacgcc tgtcaccgga tggagaagt gctctgtgc ttctgggag gccctacccc 4740  
 cctggccgccc gcccgaaga tgtggcactg tgcagagcc gccgcccgt aaggcctggg 4800  
 ggcgctgtgg tgggcccgt gccctctag agggagcagg acctcccca gacttgggt 4860  
 gggggccctg ccagctccag caccacctt gcccgaagtc ttttaacct ggtgttagca 4920  
 ttttaagag accccacagg agttctggc tgtgactaac taactgccc acccagccg 4980  
 agacctcgcc gagactgtaa ctagtgtgt ttgtacaacc aaagactcta ttttgggtt 5040



taaggagaat aaagttgact acatttttaa aaaaaaagg

5079

<210> 58

<211> 6065

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 72192179CB1

<400> 58

```

gggaactacg acctttttcag ggacttttcag aagcaacttt gttaaagcaa gaacaaaata 60
ttctaaaaag gatataaagg aaaagaatat aaacctttcaa ccaatgtggc agctttttgcc 120
tgtagagcaa gacacatcca atgtaacaga aatgaaagtc tctgaaaaaa gtcacaatac 180
atttaaggca accaacaaaa agcggggagac tgatgttcac ttgaaaagcc aggactttct 240
aatgaaaaca aatacttcca caggcttaaa aatggcaatg gaaaggctct tgaatccaat 300
caacttaacc ctgagaataa tgtaaaaagaa agtgagtgcc ccttccacc tccatctcca 360
cctcctccac caccttctaa tgcatcatct gaaattgaat ttcctcttcc tccctcacct 420
cctttgatga tgtttcctga aaaaaatggg tttcttccct cactgtccac agagaagata 480
aaggctgaat ttgaaagttt tccaggcctc cctcttccct cacctccagt agatgagaaa 540
tctgaaagag aaggttccatc gatgtttctg ccgctcctc ctccctcaac tccatctcaa 600
aagccagcac atctcctttc ctccctctgt ccggaagaag acagtggaga cttcatgcaa 660
caatattccc aaaaagaagc ctgcaactct cagaattctc aggtctaaat cataacagga 720
aaaaccggtg tgttgccacc tcccacattg cccaaaccca aacttcccaa gcatataaaa 780
gataataaga acgatttttc cccaaagttt gaactggcaa cctccctgtc agatatggaa 840
tgtaaaatta ctacctcaaa ggatcagaaa aaagtaatgg tgatgaccag cagtgaacac 900
acggagacaa agcagaacgt tattagtaag agtcttgatg aaagaaaaca attatctatt 960
gactctgcaa actgtctctc acacacagtt ccaggaactt cagcaccag gaaaaaacag 1020
attgcgctct ctataaaatc tcattcattt ccagagagtt caggacaaca aaatccaaaa 1080
ccttatatga gaaacattta agacaccttt aagtgatggc tgaagaaaaa tatagacaac 1140
aaaaagaaga aatgaaaaac agaaacagga aggttcttac tacaacatgt taaaactcaa 1200
agccaaaatc aacacataac agaggtggaa aaggaaatgc cattacaaaa aaaccaatga 1260
ggaggtttcc ctatctggaa ttgattcaga atgcactgtg gttcaaccca gccaggctc 1320
tcaaagtaat gctcgatag taggagtgtg ttctgataac caactctcca caacatcgcc 1380
agaaacagtc gctgccaaag ggtccacca tgttttagca gcttcagaag acaaagataa 1440
gatgaaaaag gaagttttac aaagctcaag ggacattatg caatccaaat cagcttgcca 1500
aattaacaaa agtcaccaag aatgtagtac ccaacaaaca caacagaaga agtatttggg 1560
gcagttgcac ttgccccaaa gcaaaccaat ttcccaaat ttcaaagtta aaaccatcaa 1620
acttccaact ctgatcata cattaatga aacagaccac agctatgaaa gtcataaaca 1680
gcaatctgag attgatgttc aaacctttac caaaaaacaa tatctgaaaa ccaagaaaac 1740
tgaagcaagc actgaatgta gtcataagca atctctggct gaaagacatt atcagttacc 1800
taagaaggag aaaagagtga cagtacaatt gcctacagaa tccatacaga agaaccagga 1860
agataagctc aagatggttc ccaggaagca aagagaattt agcggatctg acagagggaa 1920
acttccagga agtgaagaaa aaaatcaggg accatcaatg attggtcgaa aagaagagag 1980
attaataact gaaagaaaaa acgaacatct gaagaataaa tcagcaccaa aggtcgtcaa 2040
gcaaaagggt atcgatgcac atcttgattc acagactcag aattttcagc aaacacaaat 2100
acagaccgct gaaagtaaag ctgaacataa aaaattgccc cagccatata atagtctgca 2160
ggaagaaaaa tgtctcgaag tcaagggcat acaagagaaa caagtcttct ctaatactaa 2220
agattcaaag caagagatta cacagaacaa atctttcttt tctctgtgta aagaatccca 2280
gcgggatgat ggaaaagggt ccttaaatat agtggaattc ttgagaaaac gtgaagaact 2340
gcaacagatt ttgtcgagag tgaaacagtt tgaagcagag ccaaataaaa gtggccttaa 2400
aacatttcag acactattaa atactatccc aggatggctg ataagtgaag ataagagaga 2460
atatgcagtt cacattgcca tggagaataa tttagaaaaa gtaaaagaag aaataacaca 2520
tattaaaact caagcggaa atatgcttgt gtcctatgaa aatataattc agacagccat 2580
gatgtcctcc aaaacaggaa aaccgggaaa taaaactagt gaacagaaag aaaataaaat 2700
caaagtatct aatgttcatg tcagcaataa agcagctcat catgaagcaa ctgttcgtag 2760
tgccaaagag aaaacagtac agcaccaagt tgatgatagc aacattcctc ctccctcttt 2820
tcacgtgaaa acccatcagg aaattaaact cacaatagaa tctactgccc gacgaacaga 2880
aaaaacacgc ccaccgtcac caacttttat cctaataaag gacagttatg ttgaaccccc 2940
aaaccctact ccaatgtcgc aaaaatctga aattcacaga gcaaacactt ccccttctcc 3000
accaagaagg cctcttgaac aacttgtcag actcaaagac accactgcaa agttatccaa 3060

```

```

aggggccatc ccatgtccag cagcaacccc ggttccaatt gtagagaaaa ggtctgaaat 3120
catcatgtct cctgcaacac ttctgtctca aattaagata gaaactcgtg gtagggactc 3180
tccacctaca atcacaatac cagtaaatat aaatcatgct gctagtgggt ccttcagaga 3240
atctgtggac gctcaagagg aaatcaggaa agtgaggaga agagctactt atgttcataa 3300
agatggacta aattccactg atcacatggt gcccgacact gaaagttatg atgcagttga 3360
aatcatccgc aagggttgacg tgccctcctcg cctgtcagag cacacacaga gatatgaagc 3420
ggccaaccga actgttcaaa tggctgaaaa ttctgtgaat gacctgaaa atgaaataaa 3480
cagatggttc agggaaattg agcatggccc agtttctgaa gcaaagtcaa atagaagagt 3540
ttatgcaaag ggagaaacaa accataacat acaacaagaa agtcgtacat tttgtaagga 3600
ggaatttgga ttaacatctt taggaaacac gagttttaca gacttttctt gcaaacatcc 3660
tagagaactg cgagaaaaga ttctgttaa gcagcccagg atctgctctg aaaccagggtc 3720
tctaagttaa catttctcag gcatggatgc atttgagagt caaattgttg agtcgaagat 3780
gaaaacctct tcatcacata gctcagaagc tggcaaatct ggctgtgact tcaagcatgc 3840
cccaccaacc tatgaggatg tcattgctgg acatatctctg gatattctctg attcacctaa 3900
agaagtaaga aaaaattttc aaaagacgtg gcaagagagt ggaagagttt ttaaaggcct 3960
gggatatgca accgcagatg cttctgcaac tgagatgaga accacctcc aagaggaatc 4020
tgcatttata agtgaagctg ctgctccaag acaaggaaat atgtatactt tgtcaaaaga 4080
cagtttatcc aatggagtgc ctagtggcag acaagcagaa ttttcataag tctgcttcc 4140
gatgccacca ttgcaacagt aaactaagtt tgggaaatta tgcatcactt catggacaaa 4200
tatactgtaa acctcacttt aaacaacttt tcaaatccaa aggaaattat gatgaagggt 4260
ttggacataa gcagcataaa gatagatgga actgcaaaaa ccaagcaga tcagtggact 4320
ttattcctaa tgaagaacca aatatgtgta aaaatattgc agaaaacacc cttgtacctg 4380
gagatcgtaa tgaacattta gatgctggta acagtgaagg gcaaaggaat gatttgagaa 4440
aattagggga aaggggaaaa ttaaaagtca tttggcctcc ttccaaggag atccctaaga 4500
aaaccttacc ctttgaggaa gagctcaaaa tgagtaaacc taagtggcca cctgaaatga 4560
caacctgtct atccctgaa tttaaaagtg aatctctgct agaagatgtt agaactccag 4620
aaaaataagg acaaagacaa gatcactttc catttttgca gacttatcta cagtcacccc 4680
atgtttgtca gaaagaggat gttataggaa tcaaagaaat gaaaatgcct gaaggaagaa 4740
aagatgaaaa gaaggaagga aggaagaatg tgcaagatag gccgagtga gctgaagaca 4800
caaagagtaa caggaaaagt gctatggatc ttaatgacaa caataatgtg attgtgcaga 4860
gtgctgaaaa ggagaaaaat gaaaaaacta accaaactaa tgggtgcagaa gttttacagg 4920
ttactaacac tgatgatgag atgatgccag aaaatcataa agaaaatttg aataagaata 4980
ataataacaa ttatgtagca gtctcatatc tgaataattg caggcagaag acatctattt 5040
tagaatttct tgatctatta cccttgtcga gtgaagcaaa tgacactgca aatgaatag 5100
aaattgagaa gttagaaaaat acatctagaa tctcagagtt acttggtata tttgaatctg 5160
aaaagactta ttcgaggaat gtactagcaa tggctctgaa gaaacagact gacagagcag 5220
ctgctggcag tctgtgtcag cctgctccaa aaccaagcct cagcagaggg cttatggtaa 5280
agggggggaag ttcaatcatc tctcctgata caaatctctt aaacattaaa ggaagccatt 5340
caaagagcaa aaattttacac tttttctttt ctaacaccgt gaaaatcact gcattttcca 5400
agaaaaatga gaacattttc aattgtgatt taatagattc tgtagatcaa attaaaaata 5460
tgccatgctt ggattttaagg gaatttgga aggatgttaa accttggcat gttgaaacaa 5520
cagaagctgc ccgcaataat gaaaacacag gttttgatgc tctgagccat gaatgtacag 5580
ctaagccttt gtttcccaga gtggaggtgc agtcagaaca actcacggtg gaagagcaga 5640
cacttaggca ctgagagata ttgatgttct agtaaaatat ctatggccac tgacagttca 5700
tttgaggaac ttgatgtaaa catggtgttc gaaataagat tttatgaatt tggataacct 5760
ttcttgtatt acacctgttc atttttttca caatttattt tgtctatctc aatgggatat 5820
tggaatgaag agatgaaaca ctatggatat gttttccatt acatctactt ttgtttgaa 5880
tggtctgttt tctgtaaaa catcatgggt gtgattttta caaatggcac ttttagcatat 5940
attattataa cttctctgta atttctctg aaataaaatt tactgtctgt cttgttcaca 6000
caaaa gatcacctg aggtgcaaac 6060
6065

```

&lt;210&gt; 59

&lt;211&gt; 1055

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505908CB1

&lt;400&gt; 59

cgggcgacaga cggcggcagc ggcgcttgct cttggaagtt caggctcggt tgtcttttgg 60



gagccatgga gaggtagcttt tatctgcgtt actacgtggg gcacaagggc aagttcggcc 120  
acgagttcct ggagtttgag ttctgaccgg acgggaagtt aagatatgcc aacaacagca 180  
attacaagaa tgatgtcatg atcagaaaag aggagcttga aatcgtcatt ggagatgaac 240  
acattttcttt tacaacatca aaaatttggtt cccttattga tgtcaatcaa tccaaggatc 300  
cagaaggctt acgagtattt tattatcttg tccaggacct gaagtgtttg gtcttcagtc 360  
ttattggatt acacttcaag attaaaccaa tctagactga atattggtgt ggacatgggg 420  
ggtgggtggg agtagaaaaat ttgtgtata tcagggcagt atttttttat gaactataaa 480  
tgattgtctt taataaatat gtgataaaat ccaaaaaaaa aaaaaaaa 540  
aaaaaaaaa aaaaaaaa aaaaaaaa aaaaaaaa aaaaaaaa aaaaaaaa 600  
aataaaaagg gaggcacaag agttgagaca atataaacgg gggattaaat ccgggatttt 660  
ttccgggagg ggaacaaggt tcccaaagat ttccccaat gagatttggg aaacctggg 720  
aaaaaaaaag gcccaggga attttttagc cgattttaa cgcgaaaaaa ttcgggccgg 780  
aaaaaagagg ggtaaaaaa ggggtgacct aagggggggc caaaaccgaa taaattcgca 840  
aaagattggg gtgcggttta aagcggggat aacatgggg gacaatttgt ctaaaaaaga 900  
tgagcaagat aacggggaaa aacaggggtt agaaataagg ggccacacgg ggggaaagg 960  
gtagcctttt tttatgcgga aaaaggggtt aacccttaa ggccctccc agaggggtctt 1020  
taaaagggtt ccaaaggggt ttccccgat tgccg 1055

&lt;210&gt; 60

&lt;211&gt; 1651

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6590147CB1

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (1) ... (1651)

&lt;223&gt; a, t, c, g, or other

&lt;400&gt; 60

ggagagaact agtctcgcgc gcaccggctc cgcgcgccta tgagaatgag gtggtaacgg 60  
gcccccgat gaccccgct caccactgtg aggcctacag ctctgccnnn nnnnnnnnn 120  
nnnnnnnnnn nnnnnnnaag gtagctacag caagctgggt agcaggcaga tccaaaggat 180  
atcatgaagt ttccagggcc ttgggaaaac cagagattgt ctttctgtt ggaaaaggca 240  
atcactaggg aagcacagat gtggaaagtg aatgtgcgga aaatgccttc aaatcaggct 300  
catccaaat acttgagttg tattgcaatc agctgttttt tcttagctgc caagactgtt 360  
gaggaagatg agagaattcc agtactaaag gtattggcaa gagacagttt ctgtggatgt 420  
tctcatctg aaattttgag aatggagaga attattctgg ataagttgaa ttgggatctt 480  
cacacagcca caccattgga ttttcttcatt attttccatg ccattgcagt gtcaactagg 540  
cctcagttac ttttcagttt gccc aaattg agcccatctc aacatttggc agtcttacc 600  
aagcaactac ttcactgtat ggccctgcaac caacttctgc aattcagagg atccatgctt 660  
gctctggcca tggtagtct ggaatggag aaactcatc ctgattggct ttctcttaca 720  
attgaactgc ttcagaaagc acagatggat agctccagtg tgatccattg tccggagctt 780  
gtggcacatc acctttctac tctgcagtct tccctgcctc tgaattccgt ttatgtctac 840  
cgtccctca agcacaccct ggtgacctgt gacaaaggag tgttcagatt acatccctcc 900  
tctgtcccag gcccagactt ctccaaggac aacagcaagc cagaagtggc agtcagagg 960  
acagcagcct ttaccatca tctcccagct gccagtgggt gcaagcagac ctctactaaa 1020  
cgaaagtag aggaatgga agtggatgac ttctatgatg gaatcaaacg gctctataat 1080  
gaagataatg tctcagaaaa tgtgggttct gtgtgtggca ctgatttacc aagacaagag 1140  
ggacatgctt ccccttgtcc acctttgcag cctgtttctg tcatgtagtt tcaacaagt 1200  
ctacctttga gtgtaaacta aggttagacta ctttgggaat gagaacatgc aaatcagga 1260  
aaggctgtag aaggaaatat accttaacag gctgatttgg agtgagccag aaaaaaaa 1320  
taaaactctc attatttgtg tggttaatta taattcagcg ttatttaagc acataaagac 1380  
caaaaaaaa aaaaagaaat ccaaaagatc caaactttt taaaacttaa aaatctctt 1440  
tgtagtatgt cagtggccct tttttccgg ccatatggta ccgtagtgg ccccgtaaa 1500  
atttcattaa atttcgtggc agaattctgt cccctaaga tctggtaaat tccggggggc 1560  
ggggggcccc attaaagtac cgactttggc gggggcattg gtgcccgcgc agtgcaactt 1620  
tctgtgaccc gtcgtctctg tgagggggcg g 1651

&lt;210&gt; 61

<211> 2015  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 6828539CB1

<400> 61

```

cctgagcagg cagggcacct ccagggtgggc aggagctacc actcagcacc atgagcaccg 60
ccacaggccc agaagctgcc ccaaagccaa gcgccaagtc tatctatgag cagaggaagc 120
gttactccac agttgttatg gctgatgtat cccagtaccc agtcaatcac ctgggtgacgt 180
tctgcctggg tgaggacgat ggcggtgcata ccgtggagga tgcctccagg aagttggccg 240
tcatggatag ccaggggccga gtctgggcac aggagatgct gctgcgagtg tctcccagcc 300
atgtcacgct gctcgacccg gcctccaagg aggagctgga gtcgtacca ctgggcgcca 360
tcgtgcgctg tgacgcggtg atgccacccg gcaggagccg ctcggtgctg ctgctcggtg 420
gccaggaacc cgagcgcgcg cagcccagcg tgcaacttct ccaggggcctg cgccctcgggg 480
cggaagctgat ccgagaggac atccaggggg ctctgcacaa ttaccgctcg ggcccgggg 540
agcgaggggc ggcgcgctc agggccacgc aggaggagtt gcagcgcgac cgctcgccc 600
ccgctgagac ccgcccctg cagcgccgccc cgtcagtcg cgagtgatc agcaccgtag 660
agcggggccc gggccgcgga cgaccccagg cgaagcccat tccgaggca gaggaggcgc 720
agaggcctga gccggtgggg acctcgagca acgctgactc ggccctcccg gacctgggtc 780
cccggggtcc tgacctggcg gttctgcagg cggagcggga agtggacatc ctgaaccacg 840
tgctggagca ccgggaacgc ggccgcagga gccggcgccg ggcggtggg gagggttgc 900
tgacgctgcg ggccaagccg ccctcgaggg ccgagtacac cgacgtgctg cagaagatca 1020
agtacgcctt cagcctgctg gcccggtgct gcggcaacat cgccgacccc tccctctcgg 1080
agctgttgca ctctcttttc gggcctctgc agatgattgt gaacacgtcg ggggggccc 1140
agttcgcgag cagtgtgcgg cgcccgcatc tgacatcgga tgcctggcg ctgctgcggg 1200
acaacgtcac tccacgtgaa aacgagctct ggacctcgt gggggactcg tggaccgcc 1260
ccgggctgga gctgtcccgc gaggaggagc ccccatagac acccgagttc ttcagcggct 1320
gggagccgcc ggtcactgac ccgagagacc gcgcctggga ggaccagtt gagaaacagc 1380
tacagcacga gcggaggcgc cgcgaggtga cccaagcgac acagcagggc cgaggctggg 1440
aagtccgggg gcgcgcccg tccgcctggc cccgcctgac ccgactgtct tacttctac 1500
agcaaagcgc cccccaggtc gctgtcaatg gtcaccgaga cttggagcca gaatctgagc 1560
ctcagctgga gtcagagaca gcaggaaaat gggctcctgt taattatgac ttccaggccc 1620
gcaacagcag tgagctgtcg gtcaagcagc gggacgtact ggaggtcctg gaggacaaag 1680
agaaagtgtc agagctggag gcagtgatgg agaagcaaaa gaagaagggtg gaaggcgagg 1740
tggaatgga ggtcatttga cctgccaggc gcccttcgca aagagtgcag agggcccgtg 1800
ggagaacgga ctctcagac tctcccacat agcggaagtc gatcttctga aggatggcca 1860
atctgctccg gcctggtctt ccccatccc ggtggacaga cttaacgac ctgctgcag 1920
tccctccgga gaggatctgg actggctggg gcgtggagac agtctacgga 1980
aagcgcttag cagacctacg agcggggcag ttgag
2015

```

<210> 62  
 <211> 1760  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7170321CB1

<400> 62

```

gccgtgcgag ccacccgagc cgagcggggc tgggcccga aggcaagatg gtggactaca 60
gcgtgtggga ccacattgag gtgtctgatg atgaagacga gacgcacccc aacatcgaca 120
cggccagtct ctcccgctgg cggcatcagg cccgggtgga acgcatggag cagttccaga 180
aggagaagga ggaactggac aggggctgcc gcgcaagggt gccagtgcc 240
agaggaaact gaaggagctg gaggtggccg agggcggaag ggagagctg gagcgctgc 300
agggcgaggc acagcagctg cgcaaggagg agcgagctg ggagcagaag ctggaggaga 360
tgcgcaagaa ggagaagagc atgcctgga acgtggacac gctcagcaaa gacggcttca 420
gcaaggccca tgcacacgct ccgcccacct tgatgctcat ggcttcgtca ccacctccgc 480
cgtggatggg atggcgctg cgaccacggc ccgcccggcc gcgctcgagg cgctccgcag 540

```

```

ccttgcecca gccactccc cctctcacc taccacagag catggtaaat accaagcccc 600
agaagacgga ggaggactca gaggaggtga gggagcagaa acacaagacc ttcgtggaaa 660
aatacgagaa acagatcaag cactttggca tgcttcgccg ctgggatgac agccaaaagt 720
acctgtcaga caacgtccac ctggtgtgcg aggagacagc caattacctg gtcatttggg 780
gcattgacct agaggtggag gagaaatgtg cactcatgga gcaggtggcc caccagacaa 840
tcgtcatgca atttatcctg gagctggcca agagcctaaa ggtggacccc cgggcctgct 900
tccggcagtt cttactaag attaagacag ccgatcgcca gtacatggag ggcttcaacg 960
acgagctgga agccttcaag gagegtgtgc ggggcccgtgc caagctgcgc atcgagaagg 1020
ccatgaagga gtacgaggag gaggagcgca agaagcggct cggccccggc ggcctggacc 1080
ccgtcgaggt ctacgagtc ctcctgagg aactccagaa gtgcttcgat gtgaaggacg 1140
tcagatgct gcaggacgcc atcagcaaga tggaccccac cgacgcaaag taccacatgc 1200
agcgtgcat tgactctggc ctctgggtcc ccaactctaa ggccagcgag gccaaaggag 1260
gagaggaggg aggtcctggg gaccattac tggaagctgt tcccaagacg ggcgatgaga 1320
aggatgtcag tgtgtgacct gcccagcta ccaccgccac ctgcttcag gccctatgt 1380
gccccctttc agaaaacaga tagatgccat ctgcccgtc cctgacttcc tctacttgcg 1440
ctgctcgccc cagcctgggg ggcccggcca gccctccctg gcctctccac tgtctccact 1500
ctccagcgcc cattcaagtc tctgctttga gtcaaggggc ttactgcct gcagccccc 1560
atcagcatta tgccaaaggc cggggggtcc ggggaagggc agaggtcacc aggtgtgtct 1620
accaggtagt tggggagggt cccagccaa ggggcccgtc ctgctcactg ggctctgttt 1680
tactgttctg tctgtgtct gtgtcttcta tttggcaaac agcaatgatc ttccaataaa 1740
agatttcaga tgctcaaaaa

```

&lt;210&gt; 63

&lt;211&gt; 2599

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505918CB1

&lt;400&gt; 63

```

tgccggggag agcggcgcca gaagggcgcg ccgcgtcctc aagccggcac ctgagcggcg 60
gagacggctg tagcacaagg atctgcatct ccaatggata ctgagggggt tggtagctc 120
cttcagcaag ctgaacagct tgctgctgag actgagggca tctcagagct tccccatgtg 180
gaacgggaact tacaggagat ccagcaggcg ggagagcgcc tgcgttcccg taccctaaca 240
cgacgtccc aggagacggc agatgtcaag gcgtcagttc tctcggggtc tcggggactt 300
gacatatccc acatctccca gcgattggag agtctgagtg cagccaccac ctttgagcct 360
cttgagcctg tgaaggacac tgacattcag ggcttcctga agaattgaga ggacaatgcc 420
ctgtgtgtct ccacgaaga gtcccgaag aggaccttcg gcatggctga ggagtaccat 480
cgggagtc aaagcgcct tgactttact caagaaagcg agccaagcta catcagtgat 540
gcatcaggag aagacgcct tgactttact caagaaagcg agccaagcta catcagtgat 600
gtgggacccc ctggtcgaag ctctctggat aacatcgaga tggcctaagc gcggcaaat 660
tatatctata atgagaaaat tgtaaatgga cacctgcagc ctaacctggt ggacctttgt 720
gcttccgtcg cagagctgga tgataagagc atttccgaca tgtggaccat ggtaaaacaa 780
atgacagacg tgttgttgac accggcaacg gatgccctga agaaccgcag cagcgtggaa 840
gtgcgcatgg agtttgtcag gcaggccttg gctaccttg agcagagtta taagaattac 900
acccttgta ctgtctttgg aaatttgcat caggcccgagc tggcgggggt gcctgggact 960
taccaattgg ttcgaagttt cctgaacatt aaactgccag ctcccttgcc tggactacag 1020
gatggagagg tggaaaggcca tctgtgtgg gcgctaattt actactgcat gcgtgtgga 1080
gacctgcttg ccgcttcaca ggtagttaat cgagcccagc accagctggg agagttaaa 1140
acctggttcc aggagtacat gaacagcaag gacagaagat tgcctccagc tacggaaaac 1200
aagctccggc tgcattaccg tagggccctc aggaacaata accgacaacc agagtgaagt ggcggacaa 1320
gtgtactgta tcattggcag atgtgacgtc caagtgtgtt ttgacgacga tggcaccagc 1380
actgaggatt acctgtggct gaagtgaac ttccagaagc agttgttga agactatggc 1440
tcccacaag acaggctcac tctctcacag ttccttact tccaagtcc gttcctgaca 1500
gagtcacct ttacggtgaa ccagcaacce ttccttact tccaagtcc gttcctgaca 1500
gcgcagtttg aagcagcagt tgcctttctt ttcgcctgg agcggctgcg ctgccatgct 1560
gtccatgtag cactgggtgt gtttgagctg aagctgcttt taaagtctc tggacagagt 1620
gctcagctcc tcagccacga gcctgggtgac cctccttgc tgcggcggt gaactcgtg 1680
cggctcctca tgcgtgtac ccggaagttt gactccacgg acccaaggga ggcctccag 1740
tacttctatt tctcagggg tgagaaagat agtcaaggag aaaacatgtt tctgcgtgt 1800
gtgagtgagc ttgtgattga aagccgagag ttcgatatga ttcttgggaa actagagaat 1860

```

```

gacggaagta gaaagcctgg agtcatagat aagtttacta gtgacacaaa gcctattatc 1920
aacaaaagttg cttctgtggc agaaaaataa ggactgtttg aagaggcagc aaagctgtat 1980
gaccttgcca agaattgctga caaggtagtg gagctgatga acaaaactgct gagccctgtc 2040
gtccccaga tcagtgtccc gcaatccaac aaggagaggc tgaagaacat ggcaactctcc 2100
attgcccgaac ggtatagggc tcaaggaata agcgcaata aatttgtgga ctccacgttc 2160
tatcttcttt tggacttgat cacctttttt gacgagtatc atagtgggtca tattgataga 2220
gcttttgata tcattgagcg cttgaagctg gtgcccctga atcaggaaag tgtggaagag 2280
agagtggctg ccttcagaaa tttcagtgtg gaacaactcc gaagtcaagc ccgcaactctg 2340
attacctttg ctggaatgat accatacoga acgtctgggg acaccaatgc gaggtctgtg 2400
cagatggagg tcctcatgaa ttaagtgcc tgccttgggg gactctgggt cggcacactg 2460
tcagtacatc aggcacatgg gccactagg ctggggtttc tgggtttgtt tctgttgtgt 2520
tttgttttgg tttctgtatt atgtattttt gtcaacgcca ataaattttt ttgattttga 2580
tttcaaaaaa aaaaaaaaaa
2599

```

&lt;210&gt; 64

&lt;211&gt; 3305

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7505935CB1

&lt;400&gt; 64

```

cgggacagag gcagggttct gagggcaggg attccccctc gtcttggccc caccgcccgg 60
gctgggcact aaactcgggc cgcggcgggg cgagcgaggc gggctccgga gggagctgac 120
gctgatgat ggcgcagtc aacatgttta cgtggctga tgtgttgagt caagatgaac 180
tgcgcaaaaa gctataccag acgtttaagg atcgggttat actggatata ctcaagacac 240
aacttcgaaa ccagctaatt catgagttga tgcacctgt attgagtga gaactgcagc 300
ctcgggtccat ttcagtagaa gggagctccc tcttaatagg cgcctctaac tctttagtgg 360
cagatcactt acaaaagatgt ggctatgaat attcactttc tgttttcttt ccagaaagtg 420
gtttggcaaa agaaaaggta tttactatgc aggatctatt acaactcatt aaaatcaacc 480
ctacttccag tctctacaaa tctactgggtt caggatctga taaagaaaat caaaaagggtt 540
ttcttatgca ttttttaaaa gaattggcag aatatcatca agctaagag agttgttaata 600
tggaactca gacaagttcg acatttaaca gagattctct ggctgagaag cttcagctta 660
ttgatgatca gtttgcagat gcttaccctc agcgtatcaa gttcgaatct tttagaataa 720
agctaaatga gtataagaga gaaatagaag agcaacttcg ggcagaaatg tgtcaaaagt 780
tgaagttttt taaagatacc gagatagcaa aaattaaaaa ggaagcaaaa aaaaagtatg 840
aaaaggagtt aaccatgttc cagaatgatt ttgaaaaagc ttgtcaagca aaatctgaag 900
ctctcgttct tcgggaaaag agtacccttg aaagaattca caagcaccaa gagattgaaa 960
caaaagaaat ttatgctcaa aggcaacttt tactaaaaga tatggatttg ctaaggaggaa 1020
gagaagcaga gctgaagcaa agagttgaag cttttgaatt gaaccagaag ctccagggaag 1080
aaaaacataa aagcataact gaggcactta ggagacagga gcagaatata aagagttttg 1140
aggagaccta tgaccgaaag ctcaagaatg aacttctaaa gtatcaactt gaactgaagg 1200
atgactacat cattagaact aatcgactga ttgaagatga aaggaagaat aaagaaaaag 1260
ctgttcattt gcaagaggag ctcatagcta ttaattcaaa aaaggaggaa ctcaatcaat 1320
ctgtaaatcg tgtgaaagaa cttgagcttg aattagagtc tgtcaaagcc cagtcttttg 1380
caataacaaa acaaaaccat atgctgaatg aaaaggttaa agagatgagt gattattcac 1440
tactaaaaga agagaaactg gagcttctgg caaaaataa attacttaa caacaactgg 1500
aagagagtag aatgaaaac ctgctctcc taaaccgct agctcagccg gctcctgaac 1560
ttgcagtctt tcagaaagaa ctacggaag ccgaaaaggc tatagtgggt gagcatgagg 1620
agttcgaaag ctgcaggcaa gctctgcaca acaactgca agcgttctgt aaagagttta actactcagg 1740
cacagctgaa ggcccagatt ctaggttaca aagcttctgt aaagagttta actactcagg 1740
ttgccgattt aaaattgcaa ctgaagcaaa ctgagacagc cctagagaat gaagtgtact 1800
gcaatccaaa gcagtctgtg atcgatcggt ctgtcaatgg attaataaat ggcaatgtgg 1860
tgccctgcaa tggtagata agtggggatt tcttgaaaca tccttttaaa caggaaaacg 1920
ttctagcacg tatggttgca tcaaggatca caaattatcc aactgcatgg gtggagggtg 1980
gttcccctga ttctgacctt gagttttagt ccaataactaa ggcaaggggtc aaagagcttc 2040
agcaagaggc cgaacgcttg gaaaaggctt tcagaagtta ccatcgaga gtcattaaaa 2100
actctgccaa aagcccacta gcagcaaaaga gccaccatc tctgcaactg ctggaagcct 2160
tcaaaaaaat tacttcagt tccccgaaa gacatatttt tggagaggac agagttgtct 2220
ctgagcagcc tcaagtggc acactgaag aaaggaatga cgtcgtggaa gcaactgacag 2280
gcagtgcagc ctgagggtc cgcgggggca cttctccag acgctctct tccacacccc 2340

```

```

ttccaaaagc aaaaagaagc ctcgaaagtg aaatgtatct ggaaggtctg ggcagatcac 2400
acattgcttc cccagtcctt tgctctgaca gaatgcccct accatcaccc actgagtcta 2460
ggcacagcct ctccatccct cctgtctcca gccctccgga gcagaaagtg ggtctttatc 2520
gaagacaaac tgaacttcaa gacaaaagtg aattttcaga tgtggacaag ctacttttta 2580
aggataatga ggagtttgaa tcatcttttg aatgtgtaga tcagaaacaa attgaagaac 2640
aaaaggaaga agaaaaata cgggaacagc aagtgaaga acgaaggcag agagaagaaa 2700
gaaggcagag taacctacaa gaagttttag aaagggaacg aagagaacta gaaaaactgt 2760
atcaggaaag gaagatgatt gaagaatcac tgaagattaa aataaaaaag gaattagaaa 2820
tggaataatga attagaaatg agtaatcaag aaataaaaga caaatctgct cacagtgaag 2880
atccttttaga gaaatacatg aaaatcatcc agcaggagca agaccaggag tcggcagata 2940
agagctcaaa aaagatgggc caagaaggct ccctagtggg cagctgcaa tctagtgaac 3000
aagtcgaaag ttttaacaggc ttttctcatg aagaactaga cgactcttgg taacctgtt 3060
tgctgccag cttctaactt acataccgtg agaagttacg taacatttac tctttgttaa 3120
atgtttccct atcatcagac aaaactcaat aaaaatgtgt gtaatccaat gtgggttttt 3180
ttttccataa ttaattttga taccatagtg tgtgaaccaa gaataatcta gtcacgtgaa 3240
acctcttctc cagtcatagt atttctcatt cattataata aaagtaactg gcttttataa 3300
aaaaa

```

3305

&lt;210&gt; 65

&lt;211&gt; 2923

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4225965CB1

&lt;400&gt; 65

```

gggtcccca gcatcctcca cagttcctcc ggctgccttc cctaagttgc ttgggtcttgc 60
tgctacttcg gcttccccag cgtcattcta cgccccacgt gacggccctg ctctggaggcc 120
ctccagtctc tgagccagtg gcttcgggtg caggagcagg agatggaaact ggtaaaggca 180
gccctggcag aagcccttcg cctgctgcgg ctgcaggtgc ccccttctc cctgcagggc 240
tctggcacac cagctcctcc gggggacagt cttgcagccc ccccgaggact gccaccaca 300
tgcacccctt ccttggtgag ccgaggcacc cagacggaga cagaggtgga gctcaagtca 360
tccctggac cccctggcct gagcaatgga ccccagccc ctgagggggg ccagcgaaga 420
gcctagtggg acccaatctg aaggaggggg cagcagcagc agtgggtctg gctcccctgg 480
cccccgggg atcctcaggc ccttgacgac cccacagcgt gctgacacgc cgcgaagaaa 540
ttcttctctc tctcatccc cctcagagtg gcctcggcag aagctctcca ggaaggcaat 600
ctcctccgcc aacctgttag tgcggtccgg gagcacagag agccgtgggg gaaaagaccc 660
cctctccagc cctggggggc ctggatctcg gaggagcaat tacaatttgg aaggcatctc 720
agtgaagatg ttctctcgag ggcgccccat taccatgtac atcccgctc gcatcccgag 780
ccttgaggag ctgcccagtg gccacgcgc agagaccctc agccttgact ggggttatgg 840
gtacaggggt cgtgactccc gctctaattc gtttgtgttg cgctctgggg aggtggtcta 900
ctttatcgcc tgtgtggttg tgctgtaccg gcctggagga ggcccagggg gtccctggag 960
tggcggccag agacattacc gggggcacac agactgcgtt cgatgccttg ctgttcaacc 1020
tgatggtgtt cgggtagcct cgggacagac agctggagtg gataaggatg gaaagccct 1080
gcagcctgtg gttcacatct gggactcaga gacgctgttg aaactgcagg agattggact 1140
gggggccttc gagcggggtg ttggggccct ggccttttca gctgcggtac aggtgacct 1200
tctttgtgtg gtggatgatt ccaatgagca catgctgtcg gtgtgggact gcagccgggg 1260
aatgaagctg gctgagatca agagtacaaa tgactcagtc ctggccgttg gcttcaacc 1320
tcgtgacagc agctgcatcg tcaccagtgg gaaatctcac gtccacttct ggaattggag 1380
tggtggagta ggggttcctg ggaatgggac ccttaccggt aaacagggtg tctttgggaa 1440
atacaagaaa ccaagtta tcccttgctt tgtgttctt cgggatggag acattctcac 1500
tggagactca gagggaaca ttctcacctg ggggcggagc ccttcagatt ccaagacccc 1560
aggcaggggt ggcgcaaaag agacctatgg gattgtggcc caggetcacg ctcatgaagg 1620
ttctatcttc gccttgtgtc tccggagggg cgggacagtg ctgagtgttg gcgggcgagg 1680
ccgccggctg gtacagtggg ggcgggggtt ggtggccctc caggaggctg agattcccga 1740
gcacttcggg ccgtgctgag ccattgctga agggcttgge tctgagctgc tgggtgggaa 1800
cacgaagaat gcattgctga ggggagacct gggccagggc ttctcccctg taatccaggg 1860
ccacactgat gagctctggg ggctctgcac acaccctcc cagaaccgct tctcacctg 1920
cggccacgac cggcagctct gcctgtggga tggggagagc catgcaactg cctggagcat 1980
cgacctcaag gagactgggc tctgtgctga cttccaccgc agtggggcag ttgtggccgt 2040
aggactgaac acggggagggt ggttggtttt ggacacagag accagagaga tcgtgtctga 2100

```





```

tgtcattgat ggcaatgagc agctctcagt ggtccggtac agcccagatg ggttgtacct 2160
ggccatttgt tcccatgaca acgtgatcta catctatagt gtttccagtg atgggtgccc 2220
atccagccgc tttggccgct gtatgggtca ctccagcttc atcactcatc ttgactgggtc 2280
caaggatggg aatttcatca tgtccaattc tggggactat gagattcttt actgggacgt 2340
ggctggaggg tgcaagcagc tgaagaatcg ctatgagagc cgagaccggg aatgggctac 2400
ctacacctgt gtgctgggct ttcacgtcta cgtaccctgt cgctcgtgcc aaggcgccga 2460
gccgcatgta cgggggccac ggcagccacg tgaccagcgt ccgattcacg cagcagcact 2520
cgacacctgt ctgctgggc ggcaaggacg ccagcatctt ccagtggcga gtgctgggcg 2580
ctggggggcg ggggcggcg cccgccacgc cctctcgaac cccctccctg tcccccgct 2640
cctccctcga cgtttgatcg ctgctggcg ggaccgactg gcccggcgcg gtggcccccg 2700
cccgcctgc ccttccctgg cccaatcccc cagcactagg ggccgactct ttcctggact 2760
gacttcgaga cattcccgat cgcgcatttt cctggagggc gcgaacggcg cccctgcaca 2820
cactgttttag acccgctggc tgagccgggc agcccagcc taggccttga ctcccgctgc 2880
ctgctgaggg gcaataaacc agaaccaaag tcgcaaaaaa aaa 2923

```

&lt;210&gt; 66

&lt;211&gt; 11546

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7495594CB1

&lt;400&gt; 66

```

cccccttctt tcaaacattt tataggcttc agggagagag agggaggagga gagagggaag 60
aaaaaaagag gagagcgaga ggggtagaga gcgcgcgcgc tccctcccg agttcccag 120
ctgctgagga gtctggattg tgtctgtccc cagtgtcaga tgaaagggcg ctgaggtct 180
tgcccgctgc cccgcgcccc gctccgcgca cgcctctctg cgagtccggc cgccagcgc 240
ctcttcccgc ccgagccgce gcctgcgctc cggggcagcc gctctgtctc cagcgcgatg 300
tggcctcgcc tggccttttg ttgctggggt ctggcgctcg tttcgggctg ggcgaccttt 360
cagcagatgt ccccgctcg caatttcage ttcgcctct tcccagagac cgcgcccggg 420
gcccccgga gtatccccgc gccgcccgt cctggcgacg aagcggcggg gagcagagt 480
gagcggctgg gccaggcggt ccggcgacgc gtgcggctgc tgccggagct cagcagcgc 540
ctggagcttg tcttctgggt ggatgattcg tccagcgtgg gcgaagtcaa ctcccgagc 600
gagctcatgt tcgtccgcaa gctgctgtcc gacttccccg tggtgccac gccacgcgc 660
gtggccatcg tgaccttctc gtccaagaac tacgtgggtg cgcgcgtcga ttacatctcc 720
accgcgcg cgcccgagca caagtgcgcg ctgctcctcc aagagatccc tgccatctcc 780
taccgaggtg gcggcaccta caccaagggc gccttcacgc aagcgcgcga aattctctt 840
catgctagag aaaactcaac aaaagttgta tttctcatca ctgatggata ttccaatggg 900
ggagacccta gaccaattgc agcgtcactg cgagattcag gagtgagat ctactcttt 960
ggcatatggc aagggaacat tcgagagctg aatgacatgg ctccacccc aaaggaggag 1020
cactgttacc tgctacacag ttttgaagaa tttgaggctt tagctcgccg ggcattgcat 1080
gaagatctac cttctgggag ttttattcaa gatgatattg tccactgctc atatctttgt 1140
gatgaaggca aggactgctg tgaccgaatg ggaagctgca aatgtgggac acacacaggc 1200
cattttgagt gcatctgtga aaaggggtat tacgggaaag gtctgcagta tgaatgcaca 1260
gcttgcccat cggggacata caaacctgaa ggctcaccag gaggaatcag cagttgcatt 1320
ccatgtcctg atgaaaatca cacctctcca cctggaagca catcccctga agactgtgtc 1380
tgcaagagag gatacagggc atctggccag acctgtgaac ttgtccactg cctgcccctg 1440
aagcctcccg aaaatgggta ctttatccaa aacacttgca acaaccactt caatgcagcc 1500
tgtgggggtc gatgtcacc tggatttgat cttgtgggaa gcagcatcat cttatgtcta 1560
ccaatgggtt tgtggtcggg ttcagagagc tactgcagag taagaacatg tctcatctc 1620
cgccagccga aacatggcca catcagctgt tctacaagg aaatgttata taagacaaca 1680
tgtttgggtg cctgtgatga aggttacaga ctagaaggca gtgataagct tactgtcaa 1740
ggaaacagcc agtgggatgg gccagaaccc cgggtgtgtg agcgccactg ttccacctt 1800
cagatgccca aagatgtcat catatcccc cacaactgtg gcaagcagcc agccaaattt 1860
gggacgatct gctatgtaag ttgcgcgcaa gggttcattt tatctggagt caaagaaatg 1920
ctgagatgta ccacttctgg aaaatggaat gtcggagttc aggcagctgt gtgtaaagac 1980
gtggaggctc ctcaaatcaa ctgtcctaag gacatagagg ctaagactct ggaacagcaa 2040
gattctgcca atgttacctg gcagattcca acagctaaag acaactctgg tgaaaaggtg 2100
tcagtcacg ttcacccagc tttcacccca ccttaccttt tcccaattgg agatgttgt 2160
atcgtataca cggcaactga cctatccggc aaccaggcca gctgcatttt ccatatcaag 2220
gttattgatg cagaaccacc tgtcatagac tgggtgcagat ctccacctcc cgtccaggtc 2280

```

tccggagaagg	tacatgccgc	aagctgggat	gagcctcagt	tctcagacaa	ctcaggggct	2340
gaattgggtca	ttaccagaag	tcatacacaa	ggagacccttt	tccctcaagg	ggagactata	2400
gtacagtata	cagccactga	cccctcaggc	aataacagga	catgtgatat	ccatattgtc	2460
ataaaagggtt	ctccctgtga	aattccattc	acacctgtaa	atggggatatt	tatatgcact	2520
ccagataata	ctggagtcaa	ctgtacatta	acttgcttgg	agggctatga	tttcacagaa	2580
gggtctactg	acaagtatta	ttgtgcttat	gaagatggcg	tctggaaacc	aacatatacc	2640
actgaatggc	cagactgtgc	caaaaaacgt	tttgcaaacc	acgggttcaa	gtcctttgag	2700
atgtttctaca	aagcagctcg	ttgtgatgac	acagatctga	tgaagaagtt	ttctgaagca	2760
tttgagacga	ccctgggaaa	aatgggtccca	tcattttgtg	gtgatgcaga	ggacattgac	2820
tgcagactgg	aggagaacct	gaccacaaaa	tattgcctag	aatataatta	tgactatgaa	2880
aatggctttg	caattggacc	agggtggctgg	gttgacgcta	ataggctgga	ttactcttac	2940
gatgacttcc	tggacactgt	gcaagaaaca	gccacaagca	tccggcaatgc	caagtcctca	3000
cggattaaaa	gaagtgcctc	attatctgac	tataaaatta	agttaatttt	taacatcaca	3060
gctagtgtgc	cattacccga	tgaagaaat	gatacccttg	aatgggaaaa	tcagcaacga	3120
ctccttcaga	cattggaaac	tatcacaaat	aaactgaaaa	ggactctcaa	caaagacccc	3180
atgtattcct	ttcagcttgc	atcagaaata	cttatagccg	acagcaattc	attagaaaca	3240
aaaaaggctt	cccccttctg	cagaccaggc	tcagtgtctga	gagggcgtat	gtgtgtcaat	3300
tgcccttttg	gaacctatta	taatctggaa	catttcacct	gtgaaagctg	cggatcggga	3360
tcctatcaag	atgaagaagg	gcaacttgag	tgcagctttt	gcccctctgg	gatgtacacg	3420
gaatatatcc	attcaagaaa	catctctgat	tgtaaagctc	agtgtaaaca	aggcacctac	3480
tcatacagtg	gacttgagac	ttgtgaatcg	tgtccactgg	gcacttatca	gccaaaaattt	3540
ggttccccga	gctgcctctc	gtgtccagaa	aacacctcaa	ctgtgaaaag	aggagccgtg	3600
aacatttctg	catgtggagt	tccttgtcca	gaaggaaaaat	tctcgcgttc	tgggttaatg	3660
ccctgtcacc	catgtcctcg	tgactattac	caacctaatg	cagggaaggc	cttctgcctg	3720
gcctgtccct	tttatggaa	taccccatc	gctgggtcca	gatccgtcac	agaatgttca	3780
agttttagtt	caactttctc	agcggcagag	gaaagtgtgg	tgccccctgc	ctctcttgga	3840
catattaaaa	agaggcatga	aatcagcagt	caggttttcc	atgaatgctt	ctttaaccct	3900
tgccacaata	gtggaacctg	ccagcaactt	gggcgtgggt	atgtttgtct	ctgtccactt	3960
ggatatacag	gcttaaaagt	tgaacacag	atcgatgagt	gcagccact	gccttgccctc	4020
aacaatggag	tttgtaaaga	cctagtgtgg	gaattcattt	gtgagtgcct	atcagggttac	4080
acaggtcagc	ggtgtgaaga	aaatataaat	gagtgtagct	ccagtccttg	tttaataaaa	4140
ggaatctgtg	ttgatggtgt	ggctggctat	cgttgacat	gtgtgaaagg	atttgtaggc	4200
ctgcattgtg	aaacagaagt	caatgaatgc	cagtcacaa	catgcttaaa	taatgcagtc	4260
tgtgaagacc	aggttggggg	attcttgtgc	aaatgccac	ctggattttt	gggtaccoga	4320
tgtggaaaga	acgtcgatga	gtgtctcagt	cagccatgca	aaaatggagc	tacctgtaaa	4380
gacggtgcca	atagcttcag	atgcctgtgt	gcagctggct	tcacaggatc	acactgtgaa	4440
ttgaacatca	atgaatgtca	gtctaatacc	tgtagaaatc	aggccacctg	tgtggatgaa	4500
ttaaattcat	acagttgtaa	atgtcagcca	ggattttcag	gcaaaagggtg	tgaacagaa	4560
cagtctacag	gctttaacct	ggattttgaa	gtttctggca	tctatggata	tgtcatgcta	4620
gatggcatgc	tcccatctct	ccatgctcta	acctgtacct	tctggatgaa	atcctctgac	4680
gacatgaact	atggaacacc	aatctcctat	gcagttgata	acggcagcga	caataccttg	4740
ctcctgactg	attataacgg	ctgggttctt	tatgtgaatg	gcagggaaaa	gataacaaac	4800
tgtccctcgg	tgaatgatgg	cagatggcat	catattgcaa	tcacttggac	aagtgccaat	4860
ggcatctgga	aagtctatat	cgatgggaaa	ttatctgacg	gtggtgctgg	cctctctggt	4920
ggtttgccca	tacctggtgg	tgggtgcgtta	gttctggggc	aaagagcaaga	caaaaaagga	4980
gagggattca	gccagctga	gtcttttgtg	ggctccataa	gccagctcaa	cctctgggac	5040
tatgtcctgt	ctccacagca	gggtgaagtca	ctggctacct	cctgcccgaa	ggaactcagt	5100
aaaggaaacg	tgttagcatg	gectgatttc	ttgtcaggaa	ttgtggggaa	agtgaagatc	5160
gattctaaaga	gcataattttg	ttctgattgc	ccacgcttag	gagggctcagt	gcctcatctg	5220
agaactgcat	ctgaagattt	aaagccaggt	tccaaagtca	atctgttctg	tgatccaggc	5280
ttccagctgg	tccggaaacc	tgtgcagtac	tgtctgaatc	aaggacagtg	gacacaacca	5340
cttccctcact	gtgaacgcat	tagctgtggg	gtgccacctc	ctttggagaa	tggcttccat	5400
tcagccgatg	acttctatgc	tggcagcaca	gtaacctacc	agtgaacaa	tggctactat	5460
ctattgggtg	actcaaggat	gttctgtaca	gataatggga	gctggaacgg	cgtttccacca	5520
tcctgccttg	atgtcgatga	gtgtgcagtt	ggatcagatt	gtagttagca	tgttcttgc	5580
ctgaacgtag	atggatccta	catatgttca	tgtgtcccac	cgtacacagg	agatgggaaa	5640
aactgtgcag	aacctataaa	atgtaaggct	ccaggaaatc	cggaaaatgg	ccactcctca	5700
ggtgagattt	atacagtagg	tgccgcagtc	acattttcgt	gtcagggaag	ataccagttg	5760
atgggagtaa	ccaaaatcac	atgtttggag	tctggagaat	ggaatcatct	aataccatat	5820
tgtaaagctg	tttcatgtgg	taaacggct	attccagaaa	atggttgcac	tgaggagtta	5880
gcatttactt	ttggcagcaa	agtgacatat	aggtgtaata	aaggatatac	tctggccggg	5940
gataaagaat	catcctgtct	tgctaacagt	tcttgagatc	attccccctc	tgtgtgtgaa	6000
ccagtgaagt	gttctagtcc	ggaaaatata	aataatggaa	aatatatttt	gagtgggctt	6060

```

acctacacccc ctactgcac atattcatgc gatacaggat acagettaca gggcccttcc 6120
attattgaat gcacggcttc tggcatctgg gacagagcgc cacctgcctg tcacctcgctc 6180
ttctgtggag aaccacctgc catcaaagat gctgtcatta cggggaataa cttcacttttc 6240
aggaacaccg tcacttacac ttgcaaagaa ggctatactc ttgctgggtc tgacaccatt 6300
gaatgcctgg ccgacggcaa gtggagtaga agtgaccagc agtgccctggc tgtctcctgt 6360
gatgagccac ccattgtgga ccacgcctct ccagagactg cccatcggct ctttggagac 6420
attgcattct actactgctc tgatgggttac agcctagcag acaattccca gcttctctgc 6480
aatgcccagg gcaagtgggt accccagaa ggtcaagaca tgccccgttg tatagctcat 6540
ttctgtgaaa aacctccatc ggtttcctat agcatcttgg aatctgtgag caaagcaaaa 6600
tttgcagctg gctcagttgt gagctttaa tgcatggaag gctttgtact gaacacctca 6660
gcaaagattg aatgtatgag aggtgggcag tggaaacctt ccccatgtc catccagtgc 6720
atccctgtgc ggtgtggaga gccaccaagc atcatgaatg gctatgcaag tggatcaaac 6780
tacagttttg gagccatggg ggcttacagc tgcaacaagg ggttctacat caaaggggaa 6840
aagaagagca cctgcgaagc cacagggcag tggagtagtc ctataccgac gtgccaccgc 6900
gtatcttgtg gtgaaccacc taaggttgag aatggcttcc tggagcatac aactggcagg 6960
atctttgaga gtgaagttag gtatcagtggt aacctggggt ataagtcagt cggaagtctc 7020
gtatttgtct gccaaagcaa tcgccactgg cacagtgaat cccctctgat gtgtgttctc 7080
ctcgactgtg gaaaacctcc cccgatccag aatggcttca tgaaaggaga aaactttgaa 7140
gtagggtcca aggttcagtt ttctgtaat gagggttatg agcttgttgg tgacagttct 7200
tggacatgtc agaaatctgg caaatggaat aagaagtcaa atccaaagtg catgcctgcc 7260
aagtgcacag agccgcccct cttggaaaac cagctagtat taaaggagtt gaccaccag 7320
gtaggagttg tgacattttc ctgtaaagaa gggcatgtcc tgcaaggccc ctctgtcctg 7380
aaatgcttgc catccagca atggaatgac tctttccctg ttgtgaagat tgttctttgt 7440
accccaactc ccttaatttc cttgggtgct cccattcctt cttctgctct tcattttgga 7500
agtactgtca agtattcttg ttaggtggg ttttcttaa gaggaaattc taccacctc 7560
tgccaacctg atggcacctg gagctctcca ctgccagaat gtgttccagt agaattgtcc 7620
cacctgagga atcccatgg aatcattgat gtgcaaggcc ttgctatct cagcacagct 7680
ctctatacct gcaagccagg ctttgaattg gtgggaaata ctaccacct ttgtggagaa 7740
aatggtcact ggcttgagg aaaaccaaca tgtaaagcca ttgagtgcct gaaacccaag 7800
gagattttga atggcaaat ctcttacag ccctacact atggacagac cgttacctac 7860
tcttgcaacc gaggttttgc gctcgaaggt cccagtgcct tgacctgtt agagacaggt 7920
gattgggatg tagatgcccc atcttgaat gccatccact gtgattcccc acaacctatt 7980
gaaaatgggt ttgtagaagg tgcagattac agctatgggt ccataatcat ctacagtgc 8040
ttccctgggt ttcaggtggc tggcatgcc atgcagacct gtgaagagtc aggatggtca 8100
agttccatcc caacatgtat gccaatagac tgtggcctcc ctcctcatat agattttgga 8160
gactgtacta aactcaaaga tgaccagga tattttgagc aagaagacga catgatggaa 8220
gttccatatt tgactcctca cctccttat catttgggag cagtggctaa aacctgggaa 8280
aatacaaagg agtctcctgc tacacattca tcaaacttcc tgtatggtac catggtttca 8340
tacacctgta atccaggata tgaacttctg gggaaacctg tgctgatctg ccaggaagat 8400
ggaacttgga atggcagtgc accatcctgc atttcaattg aatgtgactt gcctactgct 8460
cctgaaaatg gctttttgctg ttttacagag actagcatgg gaagtgtgtt gcagtatagc 8520
tgtaaacctg gacacattct agcaggctct gacttaaggg tttgtctaga gaatagaaag 8580
tggagtgggt cctccccacg ctgtgaagcc atttcatgca aaaagccaaa tccagtcatg 8640
aatggatcca tcaaaggag caactacaca tacctgagca cgttgtacta tgagtgtgac 8700
ccggatattg tgctgaatgg cactgagagg agaactgccc aggatgacaa aaactgggat 8760
gaggatgagc ccatttgcac tectgtggac tgcagttcac cccagtctc agccaatggc 8820
caggtgagag gagacgagta cacattccaa aaagagattg aatacacttg caatgaaggg 8880
ttcttgcttg agggagccag gagtggggt tgtcttgcca atggaagttg gagtggagcc 8940
actccgact gtgtgcctgt cagatgtgcc acccgccac aactggccaa tggggtgacg 9000
gaaggcctgg actatggct catgaaggaa gtaacattcc actgtcacga gggctacatc 9060
ttgcacgggt cccaaaact cacctgtcag tcagatggca actgggatgc agagattcct 9120
ctctgtaaac cagtcaactg tggacctcct gaagatcttg cccatgggtt ccctaattgg 9180
ttttccttta tcatggggg ccatatacag tatcagtgtt ttcctgggta taagctccat 9240
ggaaattcat caagaagggt cctctccaat ggctcctgga gtggcagctc acctcctgc 9300
ctgccttgca gatgttccac accagtaatt gaatatggaa ctgtcaatgg gacagatttt 9360
gactgtggaa aggcagccc gatcagtgcc ttcaaagggt tcaagctcct aggactttct 9420
gaaatcacct gtgaagccga tggccagtgg agctctgggt tccccactg tgaacacact 9480
tcttgtgggt ctcttccaat gataccaaat gcgttcatca gtgagaccag ctcttgggag 9540
gaaaatgtga taacttacag ctgcaggtct ggatatgtca tacaaggcag ttcagatctg 9600
atttgtacag agaaaggggt atggagccag ccttatccag tctgtgagcc cttgtcctgt 9660
gggtcccccac cgtctgtcgc caatgcagtg gcaactggag aggcacccac ctatgaaagt 9720
gaagtgaac tcagatgtct ggaaggttat acgatggata cagatacaga tacattcacc 9780
tgtcagaaag atggctcgtg gttccctgag agaattcct gcagtcttaa aaaatgtcct 9840

```

ctcccgaaa	acataacaca	tatacttggt	catggggacg	atttcagtgt	gaataggcaa	9900
gtttctgtgt	catgtgcaga	aggggtatacc	tttgagggag	ttaacatata	agtatgtcag	9960
cttgatggaa	cctgggagcc	accattctcc	gatgaatctt	gcagtccagt	ttcttgtggg	10020
aaacctgaaa	gtccagaaca	tggatttggtg	gttggcagta	aatacacctt	tgaaagcaca	10080
attatttatc	agtgtgagcc	tggctatgaa	ctagagggga	acaggggaacg	tgtctgccag	10140
gagaacagac	agtggagtgg	aggggtggca	atatgcaaag	agaccagggtg	tgaaactcca	10200
cttgaatttc	tcaatgggaa	agctgacatt	gaaaacagga	cgactggacc	caacgtggta	10260
tattcctgca	acagaggcta	cagtcttgaa	gggcatctg	aggcacactg	cacagaaaat	10320
ggaacctgga	gccaccagt	cctctctgct	aaaccaaata	catgccctgt	tccttttgtg	10380
attcccgaga	atgctctgct	gtctgaaaag	gagttttatg	ttgatcagaa	tgtgtccatc	10440
aatgtaggg	aaggttttct	gctgcagggc	cacggcatca	ttacctgcaa	ccccgacgag	10500
acgtggacac	agacaagcgc	caaagtga	aaaatctcat	gtgggtccacc	agctcacgta	10560
gaaaatgcaa	ttgctcgagg	cgtaatttat	caatatggag	acatgatcac	ctactcatgt	10620
tacagtggat	acatgttgga	gggttttctg	aggagtgttt	gtttagaaaa	tggaaacatgg	10680
acatcacctc	ctatttgtag	agctgtctgt	cgattttccat	gtcagaatgg	gggcatctgc	10740
caacgcccaa	atgcttgttc	ctgtccagag	ggctggatgg	ggcgctctg	tgaagaacca	10800
atctgcattc	ttccctgtct	gaacggaggt	cgctgtgtgg	ccccttacca	gtgtgactgc	10860
ccgcctggct	ggacggggtc	tcgctgtcat	acagctgttt	gccagtctcc	ctgcttaaat	10920
ggtggaaaat	gtgtaagacc	aaaccgatgt	cactgtcttt	cttcttggac	gggacataac	10980
tgttccagga	aaaggaggac	tgggttttaa	ccactgcacg	accatctggc	tctcccaaaa	11040
gcaggatcat	ctctcctcgg	tagtgcttgg	gcacctctgga	acttatgcaa	agaaagtcca	11100
acatgggtgct	gggtcttgtt	tagtaaaactt	gttacttggg	gttacttttt	tttattttgt	11160
gatataat	gttattcctt	gtgacatact	ttcttacatg	tttccatttt	taaatatgcc	11220
tgtattttct	atataaaaaat	tatatataat	agatgctgct	ctaccctcac	aaaatgtaca	11280
tattctgctg	tctattggga	aagttcctgg	tacacat	tattcagtta	cttaaaatga	11340
tttttccatt	aaagtataatt	ttgctactaa	ataataat	gctggatagt	accattttat	11400
gaggtggcca	agggattcat	agagaggact	catgatcttt	catctgtgcc	actggcacca	11460
cataggagac	cccttacaat	aaaggagacc	cctttgttct	ggttccatct	tctatacaat	11520
atgtgtaata	ataaaagctt	tttgta				11546